UC-NRLF
B 4 031 751

17 021

UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

Vol. XVII MARCH 8, 1920 No. 28

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN NO. 4

LIBRARY OF THE UNIVERSITY OF CALIFORNIA

MENTAL TESTS FOR SCHOOL USE

BY
CHARLES E. HOLLEY

Price 50 Cents

PUBLISHED BY UNIVERSITY OF ILLINOIS
URBANA

BULLETINS OF THE BUREAU OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, *Editor*

Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

http://www.archive.org/details/mentaltestsforsc00hollrich

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN NO. 4

MENTAL TESTS FOR SCHOOL USE

BY

CHARLES E. HOLLEY, Ph.D.

Formerly Assistant Director
Bureau of Educational Research
University of Illinois

Price 50 Cents

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

COPYRIGHT, 1920
BY THE UNIVERSITY OF ILLINOIS

EDITORIAL INTRODUCTION

In the spring of 1919 Doctor Holley, while acting as assistant director of the Bureau of Educational Research of the University of Illinois, tried out six group intelligence scales in the schools of Champaign, Illinois. A kind of survey, narrow but intensive in character, was thus afforded. The data, however, with a little more analysis could be made to yield important results as to the reliabilty and validity of each of these six scales as instruments for measuring intelligence. With this thought in mind Doctor Holley carried out some of the necessary analyses and wrote the monograph which follows.

Of the six tests, three have become popular in a large way. They are the "Otis Group Intelligence Scale," the "Primer Scale," and the "Virginia Delta I" (now known as the "Intelligence Examination, Delta 2"). Besides the six which were used in this investigation there were at least three others which might have been used. In all there appear to have been nine rather well-known tests at the time the survey at Champaign was started.

Since then the number has been materially increased. Not only did several new tests come out during the school year, 1919-1920, but at least three scales, complete in every essential detail, have been published this summer in anticipation of the "fall trade." The World Book Company announces Terman's "Group Test of Mental Ability"; Lippincott annouces the "Dearborn Group Tests of Intelligence"; and the Bureau of Educational Research of the University of Illinois announces the "Illinois General Intelligence Scale." It is apparent that the movement to measure intelligence by means of group tests is well under way.

Under these circumstances school people are inquiring somewhat anxiously, "Which among all the intelligence tests is best?" Like most general questions, this has no general answer. The "best" test is the one which is most appropriate. It may not be best at all times, with all pupils, and for all purposes. The term "best" therefore needs qualification.

Nevertheless, no matter what the qualifications, there are certain characteristics which a good test—to say nothing of the best one—should possess. It should not require too much time

to administer. It should be capable of rapid and objective rating. It should correlate highly, but not too highly, with teachers' estimates of scholarship—say about 0.60. It should discriminate unmistakably between levels of intelligence which are known on other grounds to be different—e.g., the levels at different ages or grades. The subordinate exercises of which it is composed should test important mental traits and should contribute to the total score amounts proportional to the importance of these traits. Scores in the subordinate exercises should be relatively independent, for otherwise they merely tell the same story. Moreover, like the scales of which they form a part, they should discriminate between levels known to be different. Both the entire scale and its subordinate exercises should yield very few zero scores and very few scores of the highest possible value. Indeed, there should not be many scores even in the region of these extremes.

All of the scales in this investigation were examined with reference to these points. The method is of necessity largely statistical; but the outcome is practical enough. Certain very definite recommendations and suggestions are made. It is believed, therefore, that with reference to a few important tests the serious student when asking the question, "Which test is best," will find, if not a general answer, at least something fundamental and satisfactory.

B. R. BUCKINGHAM,

Director, Bureau of Educational Research,
University of Illinois.

August 26, 1920.

TABLE OF CONTENTS

PART I—THE PRESENT STATUS OF MENTAL TESTING

PAGE

A. Uses of Mental Tests 9

B. What Mental Tests Measure 16

C. Types of Mental Tests 20

PART II—COMPARISON OF GROUP MENTAL SCALES

A. Introduction 25

B. Administration of Tests 25

C. The Time Element 35

D. Comparisons of Total Scores 36

E. Analysis of the Individual Tests in the Intelligence Scales 47

F. General Comparison of Scales 62

G. Summary and Comments 65

PART III—MENTAL SURVEY OF THE CHAMPAIGN PUBLIC SCHOOLS

A. Introduction 66

B. Variations between Schools and Grades 66

C. Summary and Comments 84

Appendix 86

INDEX TO TABLES

NUMBER		PAGE
I	Total Scores in the Otis Group Intelligence Scale	26
II	Total Scores in the Theisen-Flemming Classification Test	27
III	Distribution of Scores in Pressey's Primer Scale	28
IV	Scores in Virginia Delta I.	29
V	Distribution of Scores in Sentence Vocabulary Scale	31
VI	Pupils Marking Vocabulary Sentences Correctly	33
VII	Time Consumed in Administering Scales	35
VIII	Rate of Scoring by Clerical Workers	36
IX	Correlations between Intelligence Scores and Teachers' Ratings	37
X	Teachers' Scholarship Ratings. First Grade	38
XI	Teachers' Scholarship Ratings. Fourth Grade	38
XII	Distribution of Coefficients of Correlation between Intelligence Scores and Scholarship Ratings	39
XIII	Correlations between Scales	41
XIV	Distribution of Inter-Scale Correlations	42
XV	Differences between Grade Averages with Reliabilities for Scores in Each Scale	44
XVI	Coefficient of Variability of Differences between Scores for Successive Grades	46
XVII	Discrimination between Larger Grade Ranges for Intelligence Scales	48
XVIII	Discriminative Power of Otis Scale	51
XIX	Discriminative Power of Classification Test	52
XX	Coefficients of Variability of Individual Tests for Otis Scale	54
XXI	Coefficients of Variability of Individual Tests for Classification Test	54
XXII	Coefficients of Variability of Individual Tests for Virginia Delta I.	55
XXIII	Coefficients of Variability of Individual Tests for Whipple's Group Test	55
XXIV	Coefficients of Variability of Individual Tests for Primer Scale	56
XXV	Coefficients of Variability for Vocabulary Scale	56
XXVI	Correlation between Equivalent Tests	58
XXVII	Number of Repetitions of a Test needed to Secure Reliability Indicated	60
XXVIII	Percent of Zero Scores for Different Scales	61
XXIX	General Comparisons of the Six Scales	64
XXX	School Scores for Each Scale by Grades	68
XXXI	Scores in School No. 9 for Each Scale by Rooms	69
XXXII	Medians and Quartile Deviations	70
XXXIII	Distribution of Pupils according to Their Dispersion from the Median	71
XXXIV	Distribution of Pupils in School No. 9 according to Their Dispersion from the Median	74
XXXV	Standard Scores for Indicated Ages	76
XXXVI	Distribution of Intelligence Quotients by Schools and Grades	77
XXXVII	Distribution of Intelligence Quotients in the Eighth Grade of School No. 9	79
XXXVIII	Distribution of Intelligence Quotients in the High School	79
XXXIX	Implications of Intelligence Quotients	80
XL	Intelligence Quotient Distributions for the Entire School	81
XLI	Age-Intelligence-Quotient Distribution for Sixth Grade	83
XLII	Picture Completion Test	89

MENTAL TESTS FOR SCHOOL USE

PART I—THE PRESENT STATUS OF MENTAL TESTING

A—Uses of Mental Tests

I. *The Recognition of Feeble-minded Children*—Mental ability differs among individuals from very superior to very inferior with every possible gradation between these extremes. Degrees of difference in intelligence are usually unnoticed in a community because routine life does not reveal them to the general observer. We note only the grosser variations and these, as a rule, mainly when individuals are markedly defective. The village simpleton is a familiar figure. He is the roustabout who does light chores. He is the object of ridicule, the butt of the jokes of his more intelligent associates. Among the great mass of humanity, however, discriminations are not made; and when the question of competence is raised in the school or elsewhere, there is no satisfactory basis on which agreement may rest.

In reality feeble-mindedness is present in nearly every community to a greater degree than has ever been recognized. The work of the Psychological Service of the United States Army leads one to believe that the number of people who, when mature, do not exceed the mental development of the average nine-year-old child, is probably two or three out of each hundred of the population. Some authorities have placed the dividing line between feeble-mindedness and normality between the ages of ten and eleven or eleven and twelve. On this basis the feeble-minded would probably comprise from 5 to 10 percent of the the total population of the United States.

Feeble-minded people are sadly limited in their ability to adjust themselves to social conditions. They are weak in their control of mind and body and difficult to teach. They have very poor memories and very poor discriminative power. Constant repetitions are required in order to teach them the simplest things. Many of them never learn to read or to spell or to do

simple arithmetic problems even if they are kept in school for the entire compulsory education period. This applies especially to those whose mental ability does not exceed that of the average seven-year-old child. Those whose mental development is a little better than this may do something with the ordinary schoolroom tasks but the results are hardly proportionate to the cost in time and effort. The best that can be done for these people is to train them along manual lines. They can often be taught to do the ordinary home tasks of sweeping, dusting, washing dishes, peeling potatoes, bringing in coal and wood, mowing the lawn, chopping wood, and running errands of a simple nature. Even with constant supervision, there is little hope that those whose mental rating is below seven years can be made self-supporting. Many men of eight- and nine-year mental ability, however, are getting along in the industrial world at the ordinary tasks which employ unskilled labor. They make a poor living to be sure, but they eke out an existence.

The lower types of the feeble-minded have so little mental ability that they seldom engage in crime. Occasionally feeble-minded women even of low grade become social menaces but they do not usually take the aggressive part in their misdemeanors. The higher grades of the feeble-minded, however, are a real social problem, for they are capable of participating in crimes of various sorts. Although comparatively few crimes are committed by real mental defectives, criminals actually exhibit every level of intelligence. Indeed they are more often characterized by moral than intellectual abnormality. On the other hand, people who are subnormal mentally are often model citizens, when social conduct is considered, because they have been trained to live correctly. Many criminals are defective in intelligence, but not all mental defectives are criminals even potentially.

Mental tests are of value in detecting more accurately than personal judgment the different grades of feeble-mindedness. The school may use the results of these tests in determining those for whom ordinary school work is entirely unsuited. These pupils should be given school tasks, as far as possible, that are of the manual type, because this is for them the most hopeful field of training. Even in this work the same returns should not be expected that would be secured from normal children. It is wasteful to spend a markedly disproportionate amount of the school funds on this part of the population, though a portion of

the expense may be justified on the ground that the normal children profit by the segregation of the defectives.

II. *The Recognition of Mentally Backward Children*—Above the feeble-minded in mental development come those whom we call the mentally backward. These comprise from 10 to 20 percent of the population, depending upon the criteria that are set up as the dividing lines between feeble-mindedness, backwardness, and normal development. This is the class of our population from which, as a rule, petty criminals come. These are the people who are decidedly maladjusted under present conditions and who populate our slums and hovels.

The backward learn slowly at school. They have poor memories, poor discriminative powers, and mediocre reasoning ability. If they are to be taught anything the process involves a large number of repetitions. As a rule, even when they have reached their physical maturity, they are still like children in many respects. They live in the present and care little and provide little for the future. In the schoolroom they are usually retarded; but they may have enough ability to do, in a mediocre way, the work of the grade in which they are classified if they are given extra attention. Teachers often fail to appreciate the difference between their chronological ages and those of their classmates, and, hence they fail also to detect their backwardness. Developed along some lines these backward children have instincts and emotional reactions which are those of children of their own age. This side of their nature enables them at times to surprise the teacher with what seem to be bright responses; and for this reason they are often rated higher in intelligence than they should be.

The backward children in our schools should have special treatment. If put in classes by themselves they can be given the requisite repetitions of subject-matter; and they may thus learn at the rate of which they are capable. They need a special course of study built for their needs. When in the classroom with normal children they are continuously required to do things more quickly than their mental ability permits. As a consequence they fail, although if they were given more time they could succeed. They acquire the habit of failure, of which so much has been written. Mental tests would reveal the true situation and permit proper provisions to be made.

III. *The Recognition of Normal Children who are Apparently Abnormal*—Mental tests may also be used to detect normal children who are not making the progress of which they are capable. It occasionally happens that children who have average ability fail to keep up in their school work. In such cases it would be very profitable for the teacher to take the extra time needed to coach these children in order that they may progress normally. Special attention given to normal children who have "lost out" for some reason or other, often pays very well—a point recognized by those superintendents who have organized "opportunity classes" to provide for them. Teachers of such classes, however, sometimes waste their energies on really defective children because the normal children have not been differentiated from them. If mental tests are to be used for this purpose they should be given along with tests in school subjects and, if a child is mentally normal according to the mental tests and retarded when judged by the school tests, it is obvious that extra attention given to his weakness will help to eliminate or at least to lessen it. Normal children are sometimes temperamental and fail to progress because they get "at outs" with the teacher. Situations such as these may be revealed readily and the proper remedies may then be applied.

IV. *The Discovery of Superior Children*—Mental tests have special value in the selection of superior children for special classes. These children may be just as much above the average as the backward and feeble-minded are below it. Many of them could do the ordinary work of the eight years in the elementary schools in one, two, or three years less time. There are two ways in which provision may be made for these children. One is by allowing them to skip grades now and then. This device is not to be recommended without qualifications. If a child skips a grade he misses some of the vital things at times and may be handicapped in this way. Often, however, it may be better for the markedly superior child to skip grades and thus reach his school level, than to move along in lockstep style.

The other way is to provide special classes for superior children. If a number of these children are detected by the administration of mental tests, they may be placed in a class by themselves. Under these conditions they will make much more rapid progress than they would in regular classes. Such a special class makes unusual demands on the teacher, and great care must, therefore, be taken in selecting the one who is to lead

a group of superior children. The teacher must be above the average in scholarship and be able at all times to keep up with the children in their thinking. In these special classes the work may be adjusted to the needs of the children. In some subjects they may be able to progress more rapidly than the average child. In others, the emphasis probably should be placed on supplementing the work, thus making it richer in content.

If superior children are kept in classes with normal children they often acquire bad habits. They are kept marking time at a point far below their possible working efficiency and, thus, acquire all the undesirable characteristics of mischievous children. Habits of idleness, disorder, and general inefficiency are often the result of this maladjustment. On the other hand, superior children should not be pushed too rapidly in school work as is often done when they receive extra promotions. If they are sent along at too rapid a pace they often reach levels where their mental ability is not equal to some of the tasks which are set before them. This is because the subject-matter has been graded to meet the needs of normal children whose emotional lives have matured in a definite relationship with their mental lives. Superior children with their unusual mental development are often merely normal in their emotional lives, having for example fourteen-year-old minds in ten-year-old bodies with ten-year-old emotions. When a superior child is expected to feel and think in the same terms as a child several years his chronological senior he is often unable to do so. This situation implies that if special classes become common, it may be necessary to modify the subject-matter used in classes for the gifted so that it may be fitted to them.

One argument often made against the rapid promotion of superior children is that they are soon thrown into companionship with older children. This criticism is a serious one. It is, however, anticipated by the provision for special classes advocated—a provision which groups a number of these superior children together. Where it is impossible to form classes for superior children, as will usually be the case in small school systems, one should consider the situation carefully before making extra promotions. Yet a superior child will sometimes reach a place where there is almost nothing for him to do in the grade in which he is placed. A child without something to do is a menace to himself. Under these circumstances, it may be the plain duty of the school to promote him.

V. *The Grading of Children for General Promotion*—Another use that may be made of mental tests is to reveal individual differences as a basis for grading and promotion. It is often necessary to regrade children who are changing schools. Under these conditions one cannot rely upon their marks because, coming from different schools and from different teachers, the children have been rated according to different standards. A good mental test will enable one to regrade the children in a fairly satisfactory way. These classifications can then be compared with the scholarship achievements of the pupils during the first month of the year and minor adjustments may be made. When pupils pass from one type of school to another—as from elementary to high school, or from high school to college—the application of mental tests as a basis of judging fitness for entrance and of sectioning is important. It is probable that the near future will see an extensive use of mental tests as a means of determining fitness to enter new schools.

The use of mental tests is particularly appropriate in the junior high school where sections are often formed on the basis of mental ability. In systems where a grade has four or five sections organized on this basis, it has been found that the best sections often do twice as much work as the poorest.

VI. *The Determination of General and Special Ability for Educational and Vocational Guidance*—The near future will also probably witness the extended use of mental tests in another field—the field of vocational and educational guidance. Under present conditions there are few tests which can be recommended even in a limited way as suitable for this work. This situation, however, is likely to be temporary. We are making rapid strides in the preparation of mental tests. It is probable that the year 1920 will see the publication and the standardization of a number of mental tests both general and special. Some of these, no doubt, will be suitable for this work.[1]

There are two phases of this problem. It may be attacked from the point of view of so-called general intelligence. A certain degree of mental ability is necessary for the successful negotiation of most tasks. The amount of such ability can be determined in a fairly accurate way for each vocation. Individuals who do not in this respect measure up to the minimum re-

[1] This paragraph was written in the fall of 1919. The prediction appears to have been justified.

quirement in an occupation will do well not to attempt to enter it. The same is true with respect to higher education. The tests that have been given thus far seem to show that unless a person has the necessary fundamental basis it is undesirable for him to attempt to secure a higher education. At present we are applying crude methods of selection to nearly all of these activities. Oftentimes it may be merely vague personal opinions or chance peculiarities which form the basis of the judgment of the "expert."

The other phase of this problem concerns the specific abilities which are needed in special lines of work. A few tests have been devised which attempt to pick out the mental peculiarities of people who are successful in music, art, or other specific lines. We are making a beginning in this field and probably will make rapid progress from now on. Enough has been done in industry to indicate also that different occupations make their special demands. These specific requirements can be determined and individuals, who are not equipped with the peculiar capacities needed, can be rejected by the employment office. Thus, for example, one occupation may demand clear vision, another quick perception, and a third delicate motor adjustments. The degree to which each of these traits must be present to avoid probable failure may be established, and individuals not meeting the requirements for the occupation in question may be diverted from it.

The problems of vocational and educational guidance are much more complex than the problems of the employment manager. The expert in vocational or educational guidance is expected to make a wise recommendation for every individual who comes up for an analysis. The employment manager, on the other hand, usually has a number of people from whom he is privileged to select the best. This makes it possible for the tests used by the employment manager to have an element of error in them that would be fatal to the success of the test used by the counselor of individuals. The latter is most concerned with the future possibilities of the individual. Will the boy or girl who is receiving advice develop with further education in a way that will make his or her adjustment to the required conditions easy? The future must be considered to the extent of five or ten years. On the other hand, the employment manager is concerned with the immediate present. Only rarely will he con-

sider the possibilities which may be attained by the individual five or ten years hence.

B—What Mental Tests Measure

I. *Phases of Mentality which are Measured*—In the popular mind there is much confusion as to just what mental tests measure. In general the thing sought to be measured by present mental tests is potential adaptability to conditions. How readily can the subject adjust himself to new situations? How quickly can he learn? To what extent can he profit by experience? It is a question of potential ability whether it arises from inherited native capacity or not.

From a more literal point of view, mental tests may be said to measure only the individual's performance. With his performance as a hypothesis we infer his ability. How near we come to the truth will depend upon how closely what he does corresponds to what he can do. In some cases the inference will not do the individual justice because he has not done his best—perhaps not nearly his best. But the standardization of procedure in giving tests and their repetition on different occasions with the same individuals will greatly reduce the likelihood of error in inferring ability from performance.

Moreover, we draw similar conclusions in regard to human behavior of all sorts. In other words, we infer ability from its outward manifestation in performance. A salesman's ability is gauged by the amount of his sales, a mechanic's ability by his visible product. The writer is judged by his books, the preacher by his sermons, the physician by his cures, and the business man by his holdings. In a world of action ability which does not eventuate in action is as if it were not.

Yet potential action—i.e. ability—must ever be in advance of actual performance. The margin between what can be done and what under given conditions is done varies between individuals and for the same individual at different times. When the conditions are favorable the margin is contracted and performance approaches the level of ability. Under unfavorable conditions performance may lag far behind ability.

How wide the habitual margin is for a particular individual is of little consequence. He may plead greater ability than he shows, but we shall continue to discount it to the level of his

customary performance. Indeed, we may be theoretically as well as practically correct in so doing. It may very probably be true that a person, perhaps through emotional or volitional defect, exhibits a characteristic discrepancy between intellectual ability and action—a trait which is as peculiarly his own as his blue eyes or his aquiline nose, a trait in virtue of which an unusually large amount of his mentality cannot be brought into play. We are aware that in speaking of "amount of mentality" in this connection we are using a crude expression. It is only as the mentality permits action that we can speak of its "amount." Of what is over and above that which functions, we know nothing. It may be much or little, but since it accomplishes nothing further than to provide a working margin, we may safely neglect it in our tests of intelligence.

But the variation in this margin for the same individual at different times is serious. Unless we can allow for this variation or reduce it to a negligible amount, the reliability of our results is seriously impaired. A great deal of effort has, therefore, been expended in order that our results in terms of performance may become a usable index of mental ability. It is evident that the causes of unreliability—of the variability of this margin of which we have been speaking—lie in the changing conditions under which performance takes place.

These conditions are both external and internal. The external conditions include those of temperature, ventilation, illumination, and in general all the things which may at the time be present to the senses. In mental testing a set of especially important external factors has to do with the examiner. His directions may be clear or faulty, may give too much or too little information, may give a right or wrong "mental set." His voice may be entirely or but partly audible, harsh, or pleasing. His manner may be stimulating or depressing. Effort is made to reduce the variation due to these external conditions by standardizing them. This is especially true with regard to the examiner and the directions which he is to give. Some of the other external conditions—e.g. such variations in ventilation as are commonly found in schoolrooms—do not appear to make appreciable differences in performance. On the whole we believe that variations due to external conditions have been reasonably controlled where carefully devised tests have been properly used. Greater control is possible especially through more adequate training of examiners; and through the derivation of tests which

require less special training on the part of the examiner. Progress is being made in both these directions.

Internal conditions under which performance takes place are only partly controlable. In the schoolroom we may, however, do more of this than is at first apparent. For example, the element of fatigue may be measurably controlled by a uniform schedule of work prior to the time of testing. On the other hand, remoter factors having to do with the condition in which children come to school are less easily controlled if indeed they can be controlled at all.

Unreliability due to variation in both external and internal conditions may also be reduced by repeating tests, by giving parallel tests, and by giving several different tests—in short by securing at different times additional data regarding the intelligence of the examinee. The extent to which this should be done in order to secure results of a given reliability is one of the promising statistical fields in which workers are now engaged but in which they have not, as yet, secured usable results. Meanwhile, however, it is evident that the reliability of a first determination of the mentality of an individual is greatly increased when no more than a single additional and independent determination is found to agree with it. Further determinations, if they are still substantially in agreement, will establish a degree of probability amounting to practical certainty. If determinations are not in reasonably close agreement, they may properly be regarded as chance variations from a presumably truer determination. The average of the ascertained determinations may be taken as the best representation of this truer determination.

In any event, therefore, inferring ability from performance is no new procedure. With care in administering mental tests, it is probable that we may make such inferences in reference to intelligence with reasonable accuracy.

Mental tests do not measure native capacity or general intelligence directly. They only indirectly get at these as they have been modified by experience. Even tests which are composed of the most perfect uncoachable elements are attempted more successfully by those who have had a thorough education than by those who have never been inside of a schoolroom. It is conceivable that tests may be devised which will measure pure intelligence—i.e. native capacity; but present tests are not of this nature, and it is questionable whether such tests are desira-

ble. The individuals who are being measured are not the same individuals they would have been, if they had had different experiences. The important thing from all practical points of view is the present status of the individual. Theoretically, it may be interesting to compare two people on the basis of their pure native capacity, but for most practical purposes this question is unimportant. Experience plus native capacity has made the present individual, and it is he who must be considered.

II. *Phases of Mentality which are not Measured*—Mental tests do not measure the emotional side of life. They do not test one's ability to feel or to appreciate the finer things in art, nor do they test one's feeling of respect for one's fellows. They do not measure the ability to persevere, or to "carry on," except in a very limited way. Many people with mediocre endurance have sufficient power to enable them to work at a high pitch during the brief interval of a mental test, but they would be entirely unable to work twelve hours at a stretch, day after day. Mental tests do not measure the motives which guide the conduct of an individual—his conscience, his ideals, his honesty, and dependability.

This point of view, however, does not take into account the fact that these so-called emotional characteristics are apparently correlated in general with measurable mental characteristics. Usually the most brilliant individual from the mental point of view also has a very large endowment in ideals, endurance, persistence, and appreciation. In so far as the mental and emotional characteristics of human nature are correlated, tests of mental ability are also tests of emotional ability. The exceptions, however, are responsible for much of the criticism which is directed toward mental tests.

Again, mental tests do not measure directly the ability to use habits which have been acquired. Comparatively low-grade individuals may learn to do things which are mainly habitual activities. Through much practice they may have been perfected in the habits involved and, once having learned the habits, they may be able to practice them as effectively as the average individual. Consequently, mental tests so far as they test the performance of acquired habits may not be discriminative. Special tests are needed for this purpose. Mental tests, however, will indicate to a certain degree the speed with which individuals may acquire new habits and the facility with which they may modify old habits in new situations.

C—Types of Mental Tests

I. *Individual Scales.*

a. Characteristics of individual scales—The term "individual scales" is applied to those major measuring instruments which are used to test individuals one at a time. Such scales are sometimes called interview tests. They are composed of many items of very diverse character. These different bits of test material aim to determine the stage of development in the different functions involved in mental ability. The items differ in difficulty from very easy to very hard. The difficulty of each item is known with a reasonable degree of definiteness and the response to all the questions is combined in one value This value is commonly expressed in the form of a mental age.

In general, individual tests are regarded as our most accurate instruments. They have their limitations, however. If a pupil is sick or is unusually bashful, or becomes angry, the results are not descriptive of his real ability. Anything that prevents full cooperation with the examiner will invalidate the results. The different scales which have been devised for individual use have their own special limitations. A scale which is intended to measure children only between the ages of seven and fourteen should not be used in testing the ability of people whose mental age runs above these years. Further, a test which is merely a test of performance may not be a test of linguistic or other types of ability.

All the individual scales now in use require carefully trained examiners. Each item in the test will result in accurate information only after a careful following of directions and an accurate evaluation of the responses. A slight deviation from the standard wording of the directions will materially alter the response. Leniency or severity in the scoring of responses will influence the conclusions. Each answer must be evaluated accurately in the same manner that it was evaluated when the scale was devised. This can be done only by those who have made a thorough study of the scales and have some knowledge of child psychology. Some people are by temperament entirely unfitted to give individual examinations. Being unable to secure the cooperation of the subject, they obtain erroneous results. Thus an error of as much as two or three years may be made in the determination of the mental age of the subject.

The greatest objection, however, to the individual scales is that they are time-consuming. It takes from thirty minutes to two hours to administer either of the three scales described below to a single child. This fact makes individual methods of testing so expensive that they can never come into general use in the schools.

b. Available individual scales[2]—1. "The Point Scale" by Yerkes, Bridges, and Hardwick. The manual describing this scale is published by Warwick and York, Baltimore, Maryland. The scale is composed of 20 different tests, and the total possible score is 100 points. Within each test the items are graded somewhat in difficulty. Some of the tests are much easier than others but there is no careful gradation of the tests from very easy to very difficult. When this scale was devised it was intended to be valid between the mental ages of seven and fourteen. It has been found, however, in practice that if the scale is used with adults the results are questionable when the mental age exceeds twelve. The Point Scale may be given in less time than the two following scales, but it is a comparatively inflexible instrument. It deals primarily with literary material and places the unschooled individual at a decided disadvantage. The technique of its administration is somewhat difficult and no one should attempt to use it unless he has made a careful study of the manual and has been supervised in administering it. In other words, it cannot be given by an untrained examiner. The Point Scale is a modification of the early Binet-Simon Scale with the addition of a few new elements.

2. "The Stanford Revision of the Binet-Simon Scale" by Lewis M. Terman. The manual describing this scale is entitled "The Measurement of Intelligence" by Lewis M. Terman, published by the Houghton Mifflin Co. The envelope of test-materials needed in administering the scale is furnished by the same publishers. This scale, as its name implies, is a revision and extension of the Binet-Simon Scale. It is composed of 90 different tests arranged in 12 groups corresponding to mental levels of from three to eighteen years. The large number of tests permits at least six of them to be included in each age group, thus securing a comparatively high reliability. The length of the scale, however, increases the time needed for its administration. Few workers who do thorough testing of children take less

[2] The materials needed for the administration of the three individual scales described are sold by C. H. Stoelting Co., 3047 Carroll Avenue, Chicago, Ill.

than an hour for each child when using this scale. In some cases it is necessary to use an hour and a half or two hours to complete the test according to instructions.

The scale is accompanied by very carefully prepared directions. Since it is composed of so many items, it can be administered successfully only by those who have had a thorough training. In comparison with the Point Scale it requires two or three times as much effort to learn to give the Stanford Revision. The results secured, however, are usually considered to be more significant. Individuals who have not gone to school, however, are penalized by the literary character of many of the tests and do not do themselves justice.

3. "A Scale of Performance Tests" by Rudolf Pintner and Donald G. Paterson, published by D. Appleton and Co., New York. The performance scale as devised by these authors has proved to be useful for measuring the mental ability of illiterates and foreigners. In the Psychological Service of the United States Army a modification of this scale was used with those men who could not be tested with the Point Scale or the Stanford Revision. As presented in this book, the scale is somewhat poorly adapted for school use. Modifications can be made, however, which will make it helpful in those situations where literary material cannot be used for test purposes.

II. *Group Scales.*

a. Characteristics—There are already a number of group scales for measuring the mental ability of children and adults. These are made up of several graded tests each of which is composed of individual items which are comparatively homogeneous. The theory underlying these tests is that several tests measuring different mental functions will measure general intelligence when the results are pooled. As at present arranged, it is felt that these group scales are not as accurate in their measurements as individual scales. In all probability, however, as high a degree of accuracy can be secured from the use of a number of group scales as from a single individual scale. This point has not been definitely settled at this time, however, and additional evidence is needed to guide us properly. Where a comparatively rough estimate of the mental ability of people is desired, these group scales answer the purpose very well. They have the decided advantage over the individual scales that they do not require much time per subject for their administration.

Moreover, most of them are not as complicated as the individual scales, and they can, therefore, be administered by intelligent people who have had comparatively little training. The points to be remembered in the administration of any one of the group intelligence scales are seldom as numerous as the instructions for two or three single tests in the Stanford Revision. It is very probable that group scales will supplant individual scales for general purposes. They will indicate those individuals who deviate from the norm and then, if the results need confirmation, it will be possible to give individual tests or to make additional studies of these unusual individuals.

b. Existing group scales—Nine important group intelligence scales have lately come to our attention. Doubtless there are others; for during the past six months a number of psychologists have been busy developing group scales. Under the auspices of the Bureau of Educational Research, six of these nine scales have been tried out during the past year.[3] Detailed results will be given in Part II of this report. The six scales are the following:

1. "Otis Group Intelligence Scale," devised by Dr. Arthur S. Otis; published by the World Book Company.

2. "Classification Test," devised by Dr. W. W. Theisen and Mrs. Cecile White Flemming. Announced for publication by Teachers College, Columbia University.

3. "Group Test for Grammar Grades," devised by Professor Guy M. Whipple; published by the Public School Publishing Co., Bloomington, Illinois.

4. "Primer Scale," devised and published by Mrs. Luella W. Pressey, Indiana University, Bloomington, Indiana.

5. "Virginia Delta I," devised by Professor M. E. Haggerty for the Virginia Educational Commission; published by the World Book Company under the name *Intelligence Examination, Delta 2.*

6. "Sentence Vocabulary Scale," devised by the writer; published by the Bureau of Educational Research, University of Illinois, Urbana, Illinois.

The remaining three of the nine group scales to which we have referred were not tried out. They are briefly described

[3] Spring of 1919.

below. The first had been used earlier in the year at Danville and will be reported elsewhere. The second was for more advanced pupils than we were testing, and the third became available too late for use.

1. "Indiana Group Point Scale," devised and published by Sidney L. Pressey, Indiana University, Bloomington, Indiana. This group scale was one of the first to be published. As a pioneer scale it deserves no small credit, but it contains defects in administration which will prevent it from becoming popular in its present form. It is long and very exacting on the examiner; and the scoring is somewhat difficult. It cannot be given by teachers with success unless they have been carefully trained in its administration. The units of the scale are somewhat coarse and its discrimination is not very accurate. There are ten tests, each containing 20 items which are supposed to measure ability from the third grade through the high school.

2. "Psychological Examination for College Freshmen and High School Seniors, Parts A and B," devised and published by L. L. Thurstone, Carnegie Institute of Technology, Pittsburgh, Pennsylvania. This group scale is arranged in what is known as the "Omnibus Form." Its administration is exceedingly simple since the examiner has almost nothing to do except to start and stop those taking the test. The blanks contain complete directions. Little can be said by the writer as to the value of this scale. The materials used are approximately the same as those used in the Alpha Army Test. There is little doubt but that they are difficult enough for the groups of students (college freshmen and high-school seniors) for whom they have been devised.

3. "Virginia Delta VII for Grades I to III," devised by Professor M. E. Haggerty for the Virginia Education Commission; published by the World Book Company under the name *Intelligence Examination, Delta I.* This scale is one of the latest that has come to the attention of the writer. It seems to offer possibilities which will make it valuable for the primary grades. Nothing further can be said about it at this time since no published data are available.[4]

[4] Since this was written, standards have been provided. They are furnished when the tests are purchased.

PART II—COMPARISON OF GROUP MENTAL SCALES

A—Introduction

The rapid development of group scales has always been referred to. Some were planned before the United States entered the World War, and have been gradually developed since then. Others originated in connection with the work of the Psychological Service of the United States Army. Thus a number of group scales have become available for school use without much knowledge of their appropriateness for such use. Instead of considering each of these instruments from an *a priori* point of view, we have preferred to administer them to public school children under school conditions and to draw conclusions from the facts as thus revealed.

The opportunity to do this presented itself in connection with the work of the Bureau of Educational Research during the second semester of 1918-1919. With the cooperation of teachers and supervisors, six scales were administered to the school children of Champaign, Illinois, in the elementary and high school. The following scales were used: (1) Otis Group Intelligence Scale; (2) Classification Test, Form A; (3) Whipple's Group Tests for Grammar Grades; (4) Pressey Primer Scale; (5) General Examination No. 1—Virginia Delta I, and (6) Sentence Vocabulary Scales.

B—Administration of Tests

Approximately twenty-five hundred children were tested with one or more scales. With the exception of the Sentence Vocabulary Scales, which were administered by the individual teachers, all of the scales were given by the writer or by supervisory teachers or trained workers. Consequently, all of the data, with this single exception, may be considered to have been secured by disinterested people who could be relied upon to administer the scales according to instructions.

I. *The Otis Scale*—Due to physical limitations the administration of the Otis Scale was restricted to those grades in which it was thought that it could be given with greatest success. Furthermore, it seemed best not to attempt to give it to all of the schools in the city. Consequently it was offered only to grades VI to XII inclusive. In grades VI to VII the children in one average school alone were examined. In the eighth grade

and in the high school the children were selected at random and were probably representative of these grades for the city. The first administration of the Otis Scale was in the eighth grade. The reaction of the children seemed to indicate that the scale would not be suitable for the lower intermediate grades. However, it was decided actually to give the test in the sixth and seventh grades to learn positively whether or not it was suitable for these grades. A general idea of the results obtained from administering the Otis Scale may be gathered from the central tendencies and variabilities for each grade as shown in Table I. The maximum score for this scale is 230.

TABLE I. TOTAL SCORES IN THE OTIS GROUP INTELLIGENCE SCALE

	Grade						
	VI	VII	VIII	IX	X	XI	XII
No. of pupils	47	17	61	99	95	93	54
Average	98	113	128	134	138	146	149
Standard deviation	15.5	23.5	20	25.5	25.5	24	22
Median	94	117	129	132	137	144	150

II. *Classification Test*—The number of pupils who could be tested by this scale was restricted by almost the same factors which limited the administration of the preceding scale. It seemed best not to test the same children as were tested with the Otis Scale. From the standpoint of the statistical study of the scales, it would have been desirable to measure identical children with them, but from the standpoint of the school system, it was felt to be better to test different children with each scale in order that a wider survey might result. In every case, however, at least one of the other four scales was given to each group of children examined by the Otis and Classification scales. It appeared likely that the Classification Test could not be used to advantage below the fifth grade. Accordingly, it was administered in grades V to XII inclusive, with the results indicated in Table II. In Table II are also given data obtained by Dr. Theisen from several Wisconsin communities.

TABLE II. TOTAL SCORES IN THE THEISEN-FLEMMING CLASSIFICATION TEST

	Grades							
	V	VI	VII	VIII	IX	X	XI	XII
Champaign, Ill.								
No. of pupils	81	63	62	55	61	24	33	31
Average	60	84	98	108	118	125	134	141
Standard deviation	18.5	16	18.5	16.5	22	27.5	22	26
Median	59	83	99	110	116	131	136	141
Wisconsin [a]								
No. of pupils			142	101	608	289	118	262
Average			75	90	108	115	112	123
Standard deviation			22	21	24.5	29.5	20.5	24.5
Median			74	89	109	114	112	122

[a] Unpublished material furnished through the courtesy of Dr. W. W. Theisen.

III. *Whipple's Group Test*—It did not take very much work with this scale to show its general administrative inferiority in its present form to the other scales which were studied. It was found that much time was needed to give it and that a number of things about the scoring make it unsatisfactory from that point of view. The edition of the scale used at Champaign was the first offered by Dr. Whipple. It, therefore, contained defects which have been eliminated in later editions. For example, Test 5 was not printed in a form that was intelligible to the children, and it could not be used in this study. Moreover, there was no authorized procedure by which the scores in the different tests could be converted into a total score comparable to the total scores of other scales. It was administered to but 145 children. They were in grades IV to VI inclusive.

IV. *Pressey's Primer Scale*—The Primer Scale was administered throughout the city in grades I to III inclusive. The distribution of the scores is shown in Table III.

TABLE III. DISTRIBUTION OF TOTAL SCORES IN PRESSEY'S PRIMER SCALE BY GRADES

Score	Grade		
	I	II	III
0—4	1		
5—9	2		
10—14	2		
15—19	9	4	
20—24	8	1	
25—29	8	3	1
30—34	9	3	2
35—39	16	7	5
40—44	33	11	8
45—49	18	22	8
50—54	32	29	17
55—59	13	22	21
60—64	13	31	32
65—69	6	33	31
70—74		8	27
75—79		7	23
80—84		2	10
85—89			3
90—94			1
Total . . .	170	183	189
Average . . .	43.3	56.4	64.1
Standard deviation	14	12.5	12
Median . . .	44	57	64

V. *Virginia Delta I*—The materials for this test were furnished by Dr. M. E. Haggerty, Director of the Division of Tests and Measurements, Virginia Education Commission. That norms with which to compare the results of this test, which was being given in Virginia, might be available 1,200 copies were supplied to the Bureau of Educational Research. These proved sufficient to test all the children of Champaign, Illinois, in grades III to VIII inclusive. The results are shown in Table IV.

TABLE IV. TOTAL SCORES IN VIRGINA DELTA I.

Score	Grade					
	III	IV	V	VI	VII	VIII
0— 9	1					
10— 19	3					
20— 29	11					
30— 39	24	2				
40— 49	27	22				
50— 59	27	30	19	1		
60— 69	12	45	30	8	3	1
70— 79	7	47	51	20	6	0
80— 89	1	23	36	25	11	4
90— 99	2	9	34	32	32	16
100—109		7	17	31	29	33
110—119		1	8	28	40	38
120—129		1	4	10	23	33
130—139			0	3	19	38
140—149			2	1	3	15
150—159					1	1
160—169						1
Total	115	187	201	159	167	180
Average . . .	48	69.1	82.7	102.7	111.9	117
Standard deviation	16	16	16.5	16.5	17.5	16.5
Median . . .	48	68	80	98	110	117

VI. *Sentence Vocabulary Scales.*

a. Origin—The high degree of reliability of the vocabulary test contained in the Stanford Revision of the Binet-Simon Scale suggested that this material, if arranged as a group test, might prove valuable. Accordingly, sentences were devised each of which contained one of the words from the vocabulary lists of the Stanford Revision. The last word of each sentence was one of four words placed some distance to the right of the body of the sentence. The pupil taking the test was directed to underline one of the four words in each line which completed the sentence satisfactorily. The sentences were divided into two groups of fifty each on the same basis that Dr. Terman used in dividing his list. These two groups were called Series G and Series H. When thus arranged the sentences were mimeographed and were administered to several classes of children without a time limit. This preliminary use of the material revealed merit, and it was revised to remove obvious crudities of construction. The two series were then given to all the children in grades III to XII inclusive of the Champaign public schools. The papers were scored by deducting from the number of correctly underscored words one-third the number underscored incorrectly. This was done to reduce the effect of chance. Where a child underscored more than one word in a line, the sentence was counted as omitted.

b. Results—The analysis of the results soon revealed a wide deviation for individual pupils between the scores made in Series G and in Series H, although the median and average scores for the two series were about the same in a given grade. These differences were large enough to reduce the correlations between individual scores in the two series to surprisingly low values. In no grade was the correlation over +0.58. (See Table XXVI.)

When the results of two tests of the same kind show as much deviation for the individuals as this, the obvious thing to do is to combine the two scores into single indices. This was done, and the distribution of the total scores is presented in Table V. The deviations found between grade scores in Series G and in Series H led to a rearrangement of the two series. The total number of times that each of the one hundred sentences was completed correctly was computed for each grade. These re-

sults were reduced to percents, using as a base the total number of children who took the test in the grade in question. This procedure counts errors and omissions the same. With the resulting percents as a basis, the sentences were rearranged and are now presented as Series I and Series II. It is probable that these lists are as nearly of equal difficulty for the different grades as statistical computations can insure.

TABLE V. DISTRIBUTION OF TOTAL SCORES IN THE SENTENCE VOCABULARY SCALE

Score	Total Score for Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
0— 4	17	2								
5— 9	18	1								
10—14	19	4	4							
15—19	27	10	11	2	1					
20—24	20	21	20	3	2					
25—29	15	33	32	19	2	1	1			
30—34	6	46	40	37	16	5	0			
35—39	7	21	39	28	24	11	6		3	
40—44	1	12	21	56	31	30	13	4	2	1
45—49		5	13	33	30	21	19	8	6	4
50—54		0	7	18	24	24	20	16	8	5
55—59		1	2	9	19	32	23	24	20	15
60—64		1		4	13	19	18	21	22	15
65—69				2	5	16	17	20	23	24
70—74					3	5	13	14	17	12
75—79						3	5	3	10	7
80—84							1	5	3	7
85—89								2	3	3
90—94								2	1	0
Total	130	157	189	211	168	167	136	119	118	93
Average	17.3	29	33.5	41.5	47.1	52.1	57.2	60.7	64.2	66.4
Standard-deviation	10.5	9.0	9.5	9.5	10.5	10.5	10.5	10.5	10.5	9.5
Median	21	31	33	40	47	53	56	61	64	66

Table VI shows the percent of pupils in each grade who responded correctly to each sentence. The percents are larger in most cases than they would be, if a deduction had been made for the number of times a sentence was underlined correctly by pure chance. Such a deduction was not made. If the chance factor were not present to inflate the percents, there would be an appreciable percent of pupils having zero scores. If every child had tried each sentence, the lowest percent theoretically would have been 25. It might be said in the light of the percents in this table that "a little knowledge is a dangerous thing" because the more mature high-school pupils, who tried to get some of the words by comparison of form and derivative roots, made more errors than the grade pupils who underlined purely at random.

One of the chief merits of the sentence vocabulary scale is the ease with which duplicate forms can be devised. The original 100 words contained in the Terman Vocabulary List were chosen by a random sampling method from the 1904 edition of Laird and Lee's *Vest Pocket Dictionary*. Other lists of 100 words can be selected by choosing words equally distant in the dictionary from those selected by Terman. The writer has already chosen the first and second words preceding Terman's and it is planned to present these in sentences at the first opportunity. Care should be taken in the derivation of duplicate forms to select the words in the sentences in such a manner that the ideas represented by the four completing words are approximately of the same degree of abstractness as the key-word in the sentence.

TABLE VI. PERCENT OF PUPILS MARKING VOCABULARY SENTENCES CORRECTLY

SERIES I

No. of Sentence	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	85	93	97	96	98	95	97	100	98	100
2	81	98	98	99	100	93	99	97	99	100
3	80	93	95	96	100	96	98	95	98	99
4	79	82	85	90	87	86	90	94	88	85
5	72	91	92	97	97	96	99	100	100	100
6	68	92	88	91	96	93	98	94	94	99
7	62	92	92	95	95	99	100	100	100	100
8	61	92	88	96	100	94	99	99	100	100
9	62	68	84	83	87	90	96	96	95	97
10	53	88	90	97	95	99	100	100	100	100
11	42	75	85	95	97	100	99	100	98	100
12	49	74	88	95	100	97	99	100	99	100
13	44	74	84	89	84	83	90	95	95	97
14	53	73	57	74	76	90	93	82	86	89
15	51	64	69	91	95	98	96	99	99	100
16	32	72	64	90	86	96	99	97	99	99
17	35	68	56	78	84	90	90	84	94	97
18	38	67	75	91	86	91	93	97	96	98
19	36	63	72	76	82	84	80	68	84	85
20	32	46	62	77	82	83	89	79	71	79
21	18	22	48	59	73	84	90	90	95	97
22	21	29	39	57	70	71	81	79	76	82
23	36	56	45	45	58	72	98	73	79	88
24	19	42	45	61	76	76	91	92	98	99
25	26	47	33	42	69	67	72	71	79	90
26	32	22	37	38	56	64	77	82	85	88
27	46	48	48	51	54	58	62	39	45	61
28	19	11	22	31	42	63	69	77	85	92
29	36	29	34	40	32	60	59	52	44	51
30	16	15	28	24	43	53	74	73	77	81
31	16	19	24	33	42	52	61	66	65	83
32	29	39	19	32	45	43	61	75	72	88
33	23	16	31	41	48	45	63	57	50	63
34	10	17	17	20	24	31	55	64	54	69
35	19	14	23	17	24	27	51	52	57	77
36	35	26	42	37	44	44	51	28	24	49
37	16	13	7	19	37	34	46	24	33	51
38	18	31	23	28	29	36	47	44	50	56
39	25	33	23	26	37	28	39	44	43	67
40	15	29	31	33	35	35	38	24	21	45
41	13	11	16	21	29	27	39	18	26	40
42	12	12	29	28	18	19	38	19	17	29
43	23	23	18	28	21	14	38	17	9	27
44	12	9	13	7	11	15	16	12	22	33
45	28	28	19	35	22	25	22	26	28	30
46	6	15	19	22	30	33	29	18	36	36
47	9	8	17	15	18	27	31	16	5	28
48	30	37	38	23	22	30	36	19	5	27
49	7	23	20	14	11	17	4	13	10	15
50	14	14	17	7	20	15	34	10	10	14

TABLE VI—(*Continued*)

SERIES II

No. of Sentence	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	96	98	100	99	99	89	100	100	100	100
2	81	98	96	98	95	99	100	100	100	100
3	84	89	90	94	96	100	99	98	100	100
4	77	97	93	99	100	98	99	100	99	99
5	72	95	95	97	100	97	99	100	99	100
6	68	91	86	96	100	94	99	99	98	100
7	63	79	79	89	93	92	93	94	94	99
8	60	87	86	97	96	93	97	91	98	95
9	58	82	84	95	95	97	99	98	100	100
10	58	94	88	95	97	98	99	100	98	99
11	51	77	78	82	90	94	96	98	95	96
12	59	78	69	88	95	91	97	91	99	98
13	55	72	69	87	93	94	96	91	96	98
14	48	72	81	78	88	94	99	99	100	100
15	49	72	83	91	96	93	98	100	97	95
16	35	68	75	85	90	94	99	100	98	100
17	40	65	75	79	87	93	98	94	93	97
18	41	67	67	80	77	85	92	84	90	87
19	22	58	46	74	87	80	86	83	95	95
20	27	48	66	67	82	88	86	98	95	100
21	13	35	32	63	68	66	89	71	76	73
22	35	26	40	49	67	84	80	70	79	89
23	19	41	40	57	79	84	96	96	97	98
24	12	32	32	53	67	60	82	86	88	89
25	7	12	26	34	62	79	91	85	92	92
26	33	37	39	40	59	66	88	87	88	85
27	42	41	48	42	54	49	55	46	50	55
28	6	7	13	15	45	76	79	91	98	99
29	7	17	33	26	36	55	71	66	67	86
30	14	15	26	26	47	54	64	72	75	84
31	6	19	15	31	37	52	61	76	78	81
32	16	5	13	14	44	43	61	63	61	85
33	14	26	24	24	37	42	64	57	61	63
34	18	47	26	47	48	48	57	70	85	87
35	19	14	13	13	32	24	54	40	39	54
36	16	22	20	19	37	29	53	48	38	48
37	19	19	15	25	36	24	44	30	33	53
38	23	11	16	17	26	33	46	53	61	71
39	24	32	23	19	23	27	29	30	43	47
40	20	27	25	27	41	35	51	22	22	41
41	12	5	6	4	9	8	26	32	42	63
42	32	28	35	25	34	14	40	16	19	19
43	19	25	19	17	15	20	19	29	21	35
44	14	28	20	28	24	32	32	18	24	35
45	32	26	21	29	36	20	19	16	10	30
46	10	13	10	3	46	11	25	12	16	31
47	15	33	34	43	32	35	34	23	20	29
48	11	23	22	16	21	18	22	16	5	22
49	15	32	15	22	22	17	23	22	2	17
50	12	7	15	5	4	8	5	6	10	14

C—The Time Element

One of the important considerations in selecting scales is the time needed for their administration and evaluation. Some group scales are so arranged that it takes a minimum of time to give them as well as to score them; while others go to the other extreme. Table VII presents briefly the approximate amount of time required to give those considered in this study.

TABLE VII. TIME CONSUMED IN ADMINISTERING THE SCALES

Scale	Minutes
Otis	70
Classification . . .	50
Virginia Delta I . .	30
Primer	25
Whipple's Group . .	80
Sentence Vocabulary	20—40[a]

[a] This scale is given without a time limit; but the time varies from about forty minutes in the third grade to twenty minutes or less in the twelfth.

These times are approximately those used in administering the scales at Champaign. Of course, much time can be wasted by inefficient routine methods. The papers should be distributed and collected quickly. No petty interruptions should be permitted while the pupils are working. Under these conditions the time required for the administration of the different scales will be approximately as indicated above.

After the intelligence scales have been administered much time is needed for scoring the papers. Many do not realize the tediousness of this work. It often costs more to score the papers and evaluate the results than to purchase the test materials. Data are presented here to show the approximate rate at which the scales used in this study were scored. This work was done by trained clerks who used stencils wherever possible. Few teachers will approximate this rate of work when they first attempt to score similar papers. The numbers of papers scored per hour by our clerks are indicated in Table VIII for the different scales.

TABLE VIII. RATE OF SCORING BY CLERICAL WORKERS

Scale	No. Scored Per Hour
Otis	13
Classification	15
Virginia Delta I	20
Primer	35
Whipple's Group Test	6
Sentence Vocabulary (one series)	40

The time which may be devoted to the analysis of data after papers are scored is exceedingly variable. It depends to a large extent upon the purpose to be served. However, approximately the same amount of time should be allowed for this as is needed for scoring. With the exception of Whipple's Group Test all the scales yield a total score. This fact places them on an equal basis after the total scores have been obtained. In other words, from this point on, each scale will require about the same amount of time—unless an analysis is made of some or all of the individual tests which compose the scales. This last process will generally be unimportant for school purposes.

D—Comparisons of Total Scores

I. *Correlations with Scholarship*—The correlations between the intelligence scores and scholarship are shown in Table IX. These values at first glance would seem to imply that the scales are not very reliable, that they do not adequately measure the mental characteristics important for school success. This might be the case, if the judgments of scholarship were entirely adequate. It will be worth while to consider this point briefly.

The teachers were instructed to rate the children in scholarship on a special sheet. Accompanying this sheet was a set of mimeographed instructions which directed that letter ratings should be so distributed that the teacher of a normal class would give 5 percent of the class A's, 20 percent B's, 50 percent C's, 20 percent D's, and 5 percent E's. If a group was abnormal, the teacher was asked to rate the children in comparison with all children of the same sex, race, and age. If these instructions had been carefully followed, the correlation with each scale would have been higher and more significant.

TABLE IX. CORRELATION BETWEEN SCORES IN THE INTELLIGENCE SCALES AND TEACHERS' SCHOLARSHIP RATINGS

Grade	Scale				
	Primer	Vocabulary	Va. Delta I	Classification	Otis
I	.42±.04				
II	.30±.05				
III	.30±.05	.54±.05	.57±.06		
IV		.50±.04	.45±.05		
V		.42±.03	.56±.04	.67±.03	
VI		.50±.04	.69±.03	.56±.06	.33±.07
VII		.45±.04	.71±.04	.51±.06	[a]
VIII		.46±.04	.58±.05	.28±.07	.41±.07
IX		.42±.05		.57±.06	.49±.05
X		.27±.06		.23±.13	.36±.06
XI		.59±.04		.56±.10	.55±.05
XII		.53±.04		.64±.10	.41±.07

[a] Cases were too few to be significant.

The teachers, however, were not able to follow the instructions very closely. This fact is shown by Tables X and XI which present the distributions of the scholarship rating for the first and fourth grades respectively, these grades having been taken as typical. Among the different schools it is evident even without converting the number of ratings into percents that there are wide deviations from the suggested percentage distribution of rating. Even when the ratings for all the schools are combined and converted into percents the discrepancy between the actual and theoretical distribution is still evident. It is clear, for example, that the first- and fourth-grade teachers gave a great many more A's than would have been expected. The number of C's was appreciably below the standard number, while the numbers of B's and D's (at least in the first grade) were of about the right order of magnitude.

The average grade for each school was computed by allowing the customary ratings of 5, 4, 3, 2, and 1 respectively for the letters A, B, C, D, and E. These averages conceal a great deal. A teacher for example, may give too many A's but if she balances them by giving too many E's the average may turn out to be 3 and the impression may be created that the distribution was correct. Nevertheless the averages do serve to indicate whether there is a constant bias on the part of the teacher in question in virtue of which she rates everybody too high or too low. There is, in the first grade, a slight tendency for teachers to rate their children above 3—that is above the expected aver-

age. This tendency is considerably more marked in the fourth grade. These deviations from the expected average are not at all accounted for by the scores in the mental tests at the schools in question. Median scores in the Primer Scale are shown in Table X and those for the Vocabulary Scale are shown in Table XI.

TABLE X. DISTRIBUTION OF TEACHERS' SCHOLARSHIP RATINGS FIRST GRADE

Rating	Number of Ratings at School:						Ratings at all Schools	
	1	2	3	4	5	6	No.	Percent
A	2	6	8	6	4	7	33	17
B	3	14	5	0	7	7	36	19
C	17	7	8	14	11	7	64	33
D	8	7	8	8	6	1	38	20
E	3	4	4	3	4	3	21	11
Average	2.8	3.3	3.1	3	3	3.5	3.1	
Median Score Primer Scale	50	48	40	43	41	46	44	

TABLE XI. DISTRIBUTION OF TEACHERS' SCHOLARSHIP RATINGS FOURTH GRADE

Rating	Number of Ratings at School:						Ratings at all Schools	
	1	2	3	4	5	6	No.	Percent
A	10	6	13	4	0	1	34	18
B	3	8	16	17	5	4	53	29
C	19	14	8	10	13	5	69	37
D	5	7	3	2	3	2	22	12
E	0	1	0	0	6	1	8	4
Average	3.5	3.3	3.9	3.7	2.6	3	3.4	
Median Score Vocab. Scale	29	33	30	32	28	30	32	

The effect of these individual variations on the scholarship ratings by the teachers is to lower the coefficients of correlation. But another factor which diminishes the correlations has to do with effort. It is well known that many school children are working far below the limit of their ability. If a superior child does not apply himself his scholarship rating may be mediocre or even poor, although his intelligence score may be high. Indeed, the novelty of the test situation and the shortness

of the effort required frequently combine to secure a performance which corresponds more closely to actual ability than does the sustained routine performance of the classroom. Again, the exceptional industry of a child of ordinary ability may place him among the best, or at least much above his companions of equal general ability. Other factors, such as sickness, irregular attendance, and change of schools may cause a child's scholarship to be rated much below what it would be under normal conditions.

Under these circumstances, the distribution of the coefficients of correlation for the different tests reveal a reasonably high correspondence with scholarship ratings. (See Table XII.) The average for the entire group is +0.462. If the correlations for the Primer Scale are omitted, the average is a little higher, namely, +0.497.

TABLE XII. DISTRIBUTION OF COEFFICIENTS OF CORRELATION BETWEEN INTELLIGENCE SCORES AND SCHOLARSHIP RATINGS

Coefficients	Number
.70—.79	1
.60—.69	4
.50—.59	12
.40—.49	9
.30—.39	4
.20—.29	3
Average	.462

It was thought that it might be significant to combine the scores of two somewhat dissimilar scales like the Virginia Delta I and the Sentence Vocabulary.[5] The combined scores of these two scales ought to show a higher correlation with scholarship then either alone, if they measure different phases of intelligence accurately, and to the extent that the ratings of scholarship are reliable. The resulting coefficients of correlation are as follows: Grade III, 0.64; Grade IV, 0.44; Grade V, 0.52; Grade VI, 0.54; Grade VII, 0.52; and Grade VIII, 0.54; Average 0.53. These figures do not show that the combination has very materially raised the correlations.

Our evidence as well as that exhibited in other investi-

[5] The scores were combined by dropping the score made in Test 3 of the Virginia Delta I, multiplying the Sentence Vocabulary score by two and finding the total. This procedure attempted to give equal weight to both of the scales. Test 3 was dropped because it is not discriminative.

gations tends to show that while scholarship and intelligence are by no means independent, their relationship is also by no means perfect. Each is affected by conditions which do not affect the other to the same degree. Even if scholarship is accurately judged by teachers, perfect correlation cannot be expected. It is probable that a coefficient of much more than +0.60 between mental test results and estimates of scholarship would mean either that the test or the estimates were faulty. The test might be such that success in it depended too much upon schooling; or the estimates of scholarship might be too greatly influenced by the notion of natural ability.

II. *Correlations between the Scales*—The correlations between the different scales administered in this study are interesting and suggestive. As many as the data permit are shown in Table XIII. In some cases results from the same grade in several schools were used; in others the correlations had to be determined for the grade of one school only. Determining these values for a single class applies a much more rigid standard to a scale than would be the case, if correlations were computed from the combined results for different classes in the same grade or for different grades. Note, for example, that the correlations of the Sentence Vocabulary Scale with the Virginia Scale Delta I are without exception higher for all schools (column 2) than they are for one school (column 3). This is significant. In all probability it implies that the true correlation between the Sentence Vocabulary Scale and Virginia Delta I is appreciably higher than is here indicated.

In general, higher correlations are found where results from several different grades are used. This is because such a selection gives a greater spread of abilities. The two combination correlations presented in Table XIII for the Otis Scale reveal this tendency. (See the last entries in columns 5 and 7.) This procedure has sometimes been adopted in studies of mental tests. But a high correlation of this sort is not so significant as a high correlation secured from the more homogeneous material of a single grade. Since intelligence scales will be most useful, if they distinguish between the children of a single grade, this rigid test will be employed in the consideration of the merits of the different scales. This criterion should not be confused, however, with the procedure of others who compute correlations by using data from the combination of several grades.

TABLE XIII. CORRELATIONS BETWEEN SCALES [a]

GRADE	SENTENCE VOCABULARY WITH:					VIRGINIA DELTA I WITH:	
	Primer	Virginia Delta I		Classification	Otis	Classification	Otis
		All Schools	School 1				
	1	2	3	4	5	6	7
III	.40±.06	.89±.02	.86±.03				
IV		.60±.04	.53±.09				
V		.66±.05	.59±.09	.52±.06		.69±.04	
VI		.62±.02	.60±.09	.51±.05	.64±.06	.72±.04	.68±.06
VII		.78±.04	.73±.06	.33±.07	[b]	.81±.03	[b]
VIII		.69±.03		.76±.04	.67±.04	.65±.04	.89±.02
IX				.78±.04	.72±.03		
X				.37±.09	.53±.09		
XI				.67±.05	.77±.03		
XII				.62±.10	.62±.06		
Average		.71	.66	.57	.66	.72	.78
VI, VII, VIII					.79±.02		.82±.02

[a] No correlations are given for Otis with Classification nor for Virginia Delta I with Primer because in neither case did the same group of pupils take both tests.

[b] Too few pupils took these tests to warrant calculating the coefficients.

The coefficients presented in Table XIII are significantly high. If all of the correlations (except those of column 3 and those for combined grades) are distributed, they are as shown in Table XIV.

TABLE XIV. DISTRIBUTION OF INTER-SCALE CORRELATIONS

CORRELATIONS	ALL TESTS
.80—.89	3
.70—.79	6
.60—.69	12
.50—.59	3
.40—.49	1
.30—.39	2

The high correlation values that have been obtained where two scales have been given to the children in a grade seem to indicate that these scales in the main are measuring much the same thing. When the diverse character of the measuring instruments is considered this fact is somewhat remarkable. Of course, the values are all low enough to lead one to be somewhat conservative in drawing conclusions concerning individual pupils from a single test.

III. *Reliability of Total Scores*—The correlation discussed in the two preceding sections may be considered to indicate the worth of mental tests from the point of view of relationships. These correlations, however, do not reveal many of the things which one would like to know about the different intelligence scales. It may be a valuable thing to have tests which correlate highly with one another or it may be the reverse. Further, it may be worth while to have tests which correspond closely to the scholarship ratings made by teachers; but on the other hand, if the scholarship ratings, given by the teachers under the conditions described above, do not forecast the real possibilities of pupils, high correlation with such scholarship ratings may not be either desirable or informing. It is conceivable that mental scales may be devised which to a greater extent than is true of these scales will direct attention to pupils who are brilliant, average, or mediocre, in a way that will enable teachers to develop their talents. It may be that the methods of education now in use are pedantic or that intelligence scales measure qualities which are highly desirable in life, but which do not function in school work. These questions cannot be settled here; but it will be worth while to consider the scales from another point of view, namely that of their power to discriminate between different intelligence levels.

Much of the discussion from this point on will make the assumption that there is a difference in intelligence level be-

tween grades—that third-grade children are, in general, more mature in intelligence than second-grade children, that fourth-grade children are still more mature than third-grade children, and so on. As the higher grades are reached, this difference probably decreases in absolute value. Correspondingly, the grade intervals shown by a test at different grade levels should probably not be equal. The differences, however, are taken to be appreciable; and good tests ought to reveal them clearly. Further, it should be possible to magnify these amounts somewhat by the use of scales especially devised to discriminate at the higher ranges. This assumption is approximately that of Binet, Terman, and others who developed the individual intelligence scales. There is this difference, however, the earlier workers made their groupings upon an age basis. The groupings here made are upon the school-grade basis. This basis was used, first, because the age groups are not completely represented in school. The least intelligent children of ages five, six, and seven have sometimes not entered school, while the more brilliant ones beyond the age of fifteen have in many instances, completed the public school. Consequently, we can use twelve unselected grade groups, while we would not have that number of comparable age groups. The grade basis was used, second, because it is more serviceable. In the schools children are classified by grades, not by ages. If a test is given, it is given to the children of the same grade, not to those of the same age. Standards for the grades and differences between grades are therefore more immediately useful than similar facts on the basis of age.

With the assumption made in the preceding paragraph we may compare the different intelligence scales on the basis of their power to discriminate between grades. If the fourth grade is more mature than the third there should be a difference between the third grade and the fourth grade in intelligence scores. That scale which reveals the most reliable differences between averages for the different grades may be considered to be the most discriminative. These facts have been computed for the different scales and are presented in Table XV.

TABLE XV. DIFFERENCES BETWEEN GRADE AVERAGES WITH RELIABILITIES (PROBABLE ERRORS) FOR TOTAL SCORES IN EACH SCALE

Grades	Differences between Averages						
	Primer	Vocabulary	Otis	Classification		Virginia	Virginia & Vocabulary
				Wisconsin	Champaign		
I & II	13.1±0.9						
II & III	7.7±0.9						
III & IV	.....	11.7±0.8	.....	.....	.....	21.1±1.2	39.7±3.3
IV & V	.....	4.5±0.7	.....	.....	.....	13.6±1.1	19.4±2.3
V & VI	.....	8.0±0.6	.....	.....	23.7±1.9	20.0±1.2	28.5±2.3
VI & VII	.....	5.6±0.7	15.2±4.1	.....	13.5±2.1	9.2±1.2	24.2±2.6
VII & VIII	.....	5.0±0.7	15.0±4.1	15.3±1.7	10.7±2.1	5.1±1.2	17.7±2.8
VIII & IX	.....	5.1±0.8	5.5±2.5	17.5±1.6	10.1±2.4		
IX & X	.....	3.5±0.9	4.7±2.5	7.3±1.3	6.2±3.6		
X & XI	.....	3.5±0.9	7.3±2.5	−2.4±1.7	9.6±4.0		
XI & XII	.....	2.2±0.9	3.3±2.6	10.2±1.6	7.3±4.0		

Since the different scales involve different numbers of units, a unit of one (e.g. the Vocabulary Scale with a possible score of 100) is not equivalent to a unit of another (e.g., the Otis Scale with a possible score of 230). The differences, however, between the grade averages shown in Table XV may be reduced to the same basis, either by dividing the grade differences by their probable errors (thus making the differences comparable), or reciprocally, by dividing the probable errors by the differences. This latter procedure gives us the so-called coefficient of variability. It is commonly held in statistical circles that a quantity should be at least three times its probable error to be worthy of statistical consideration, or, in other words that the probable error should be no more than one-third of the quantity.

With this standard we may examine the coefficients of variability presented in Table XVI remembering that the smaller they are the more reliable are the grade differences given in Table XV and that according to the standard they should be less than 0.333. Comparing the values grade by grade, it can be seen readily that some of the scales are much more discriminative than others. The best values for grades III to VI are shown by the Virginia Scale. From grades VI to XII the Vocabulary Scale seems to show the best discrimination. It is somewhat unfair to compare the scores secured in Champaign with the results furnished by Dr. Theisen for the Classification Test in Wisconsin, for it is altogether possible that a measurement of the children in Wisconsin with the Otis and Vocabulary scales would reveal a greater degree of discrimination than these tests revealed in Champaign. However, the comparatively good values shown by the Wisconsin figures are interesting and suggestive. To a certain extent the large (and hence unfavorable) coefficients yielded for the Otis and Classification scales by the Champaign data are due to the fact that the number of children tested was less than 100 in every grade. But this is not the sole reason. The numbers of children who took these tests were no larger in the grades below the high school, yet the coefficients of variability are much smaller for those grades. In the high-school grades only one of the eight coefficients of variability for the Otis and Classification scales in Champaign is less than 0.333. This would imply, so far as our data permit an inference, that neither test is as reliable for high-school work as one would wish.

TABLE XVI. COEFFICIENT OF VARIABILITY OF DIFFERENCES BETWEEN TOTAL SCORES FOR SUCCESSIVE GRADES

Grades	Primer	Vocabulary	Otis	Classification		Virginia	Virginia & Vocabulary
				Wisconsin	Champaign		
I & II	0.068						
II & III	0.117						
III & IV	.....	0.068	.....	.....	.....	0.056	0.083
IV & V	.....	0.155	.....	.....	.....	0.108	0.118
V & VI	.....	0.075	.....	.....	0.080	0.060	0.080
VI & VII	.....	0.125	0.269	.....	0.155	0.130	0.107
VII & VIII	.....	0.140	0.273	0.111	0.196	0.235	0.158
VIII & IX	.....	0.156	0.454	0.091	0.237		
IX & X	.....	0.257	0.531	0.178	0.580		
X & XI	.....	0.257	0.342	0.708	0.416		
XI & XII	.....	0.409	0.787	0.156	0.547		

The discriminative power revealed by the scales for wider ranges of grades is also interesting. A few of the many possible facts of this sort are presented in Table XVII. Observe that the grades set up in this table have reference to the elementary school, the junior high school, and the senior high school. The greatest discriminative ability is shown by the sum of the Virginia and Vocabulary scales. The Virginia Scale by itself seems to be twice as discriminating as the Vocabulary Scale for the elementary school. When the Vocabulary, Otis, and Classification scales are compared for grades VI to IX, they rank in that order. For the last three years of the high school the Classification Test is first, the Vocabulary second, and the Otis third. These facts imply not only that some of the scales are more discriminative than others, but also that we may secure better scales than any we now have.

E—Analysis of the Individual Tests in the Intelligence Scales

I. *Individual Test Scores*—Each of the instruments used in this investigation for measuring intelligence (except the Sentence Vocabulary) consists of several series of questions or things to do. Each of these instruments without regard to its actual title may be called a *scale;* and each separately organized group of questions and things to do may be called a *test.* Thus, the Otis Group Intelligence Scale has ten tests; Whipple's Group Test for Grammar Grades has six tests; the Virginia Delta I has six tests; the Classification Test has eight tests; and the Primer Scale has four tests. Since the total score on a scale is to be taken as indicative of intelligence, the theory is that each of the component tests shall "tap" important elements of intelligence, and that the score on each test shall enter into the total for the scale to a degree that will give the proper emphasis to the element or phase of intelligence to which the test relates.

TABLE XVII. DISCRIMINATION BETWEEN LARGER GRADE RANGES FOR INTELLIGENCE SCALES

Scale	Grades III to VIII		Grades VI to XI		Grades IX to XII	
	Difference between Averages	Coefficient of Variability	Difference between Averages	Coefficient of Variability	Difference between Averages	Coefficient of Variability
Vocabulary	34.8±1.2	0.034	15.7±0.7	0.044	9.2±0.9	0.098
Virginia	69.0±1.3	0.019				
Virginia and Vocabulary	129.5±0.3	0.003				
Otis			35.7±2.3	0.064	15.3±2.1	0.137
Classification:						
Champaign			34.3±2.3	0.067	23.1±2.1	0.091
Wisconsin					15.1±1.2	0.079

The tests of which each scale is composed are as follows:

Otis Group Intelligence Scale

I. Following Directions
II. Opposites
III. Disarranged Sentences
IV. Proverbs
V. Arithmetic
VI. Geometric Figures
VII. Analogies
VIII. Similarities
IX. Narrative Completion
X. Memory

Classification Test

I. Following Directions
II. Synonym-Antonym
III. Arithmetic
IV. Common Sense
V. Completion
VI. Analogies
VII. Number Completion
VIII. Information

Virginia Delta I

I. True-False
II. Arithmetic
III. Picture Completion
IV. Synonym-Antonym
V. Common Sense
VI. Information

Whipple's Group Test for Grammar Grades

I. Arithmetic
II. Completion
III. Substitution
IV. Reasoning
Part I
Part II
Part III
V. Punched-Hole Test
VI. Proverbs

Pressey Primer Scale

I. Dot-Pattern
II. Classification
III. Form Board
IV. Absurdities

Now, it will probably occur to anyone who has considered the foregoing material thoughtfully that the various tests which compose the different scales are not likely to be of the same value. To determine the extent to which this is the case, distributions were made for each test by grades. These distributions and interpretations were enlightening, but the limitations of this bulletin permit the presentation of no more than the general features.

A marked difference in the curves of distribution was shown by the several tests. Differences were also shown when the school-grade distributions were compared with each other for tests of the same kind occurring in different scales. In other words, some of the tests were not suited to the grades in which they were used, were too hard or too easy or too irregular. These facts will be presented more in detail in the next section.

II. *Differences betwen Successive Grade Averages*—The differences between the grade averages for each of the tests were computed. These revealed wide divergencies, showing that some of the tests were poorly adapted to the work that was expected of them.

The Otis Scale contains tests which are of small diagnostic value when discrimination between successive mental levels is sought. A number of inversions, cases where the score in a test decreased with the next higher grade, were noticed. While these might be due to some degree to an insufficiency of cases, this would not fully explain the fact that some tests showed inversions while others did not. The inversions suggest poorly constructed tests. For each test of the Otis Scale the differences between the average score in the highest and lowest grade was determined. It was thus found that the amounts contributed to the difference between the scores for the entire scale varied markedly, being seven times as large for Test IV (Proverbs) and Test IX (Narrative Completion) as for Test VIII (Similarities). The amount of this difference for each test—i.e., of the difference between the average score of

the lowest and highest grades concerned—may be taken as one indication of the discriminative power of the test. The tests in the Otis Scale arranged in order from the most discriminative to the least according to the differences between the sixth and twelfth-grade averages are given in Table XVIII. As a whole, the Otis Scale showed more discriminative power in the grammar grades than in the high school.

TABLE XVIII. DISCRIMINATIVE POWER OF TESTS. OTIS SCALE

Test	Difference between Sixth and Twelfth Grade Averages	Probable Error of This Difference
IV	8.5	0.5
IX	8.4	0.7
III	8.2	0.5
II	6.7	0.5
VII	4.3	0.5
VI	4.0	0.4
V	2.6	0.4
X	2.0	0.4
I	1.8	0.4
VIII	1.2	0.4

The Classification Test is better organized than the Otis Scale when the power of the individual tests to discriminate between successive grades is used as a criterion. Although no more of the Classification Test were administered than of the Otis, there were in the former case fewer negative differences between the averages of successive grades.

Moreover, judged by the larger recorded differences between the performance of sixth- and twelfth-grade children, the component tests of the Classification Test are considerably were found as shown in Table XIX. Again, taken as a whole, this scale, like the Otis, is more discriminative in the grammar grades than in the high school.

TABLE XIX. DISCRIMINATIVE POWER OF TESTS. CLASSIFICATION TEST

Test.	Difference between Sixth and Twelfth Grade Averages	Probable Error of This Difference
VIII	12.2	0.6
II	11.9	1.0
VI	9.9	1.0
V	7.2	0.5
III	6.3	0.4
I	5.0	0.3
VII	3.8	0.4
IV	3.6	0.3

Marked differences between the discrimination of tests of the Virginia Delta I Scale are also evident. Indeed, the incomplete picture (Test III) is apparently not suitable above the primary grades. Arranged in the order of the differences between averages for grades III to VIII the tests are as follows: Test VI, 21.9 ± 0.4; Test I, 14.8 ± 0.5; Test IV, 14.7 ± 0.4; Test V, 8.2 ± 0.1; Test II, 6.3 ± 0.1; and Test III, 1.1 ± 0.2.

There is very little difference in discrimination between the tests of the Pressey Primer Scale or between the two vocabulary scales. Such differences as exist do not materially affect the use of either of these measuring instruments.

In general it may be said that the present scales have been arranged without a careful analysis of their component parts. They are poorly balanced when the individual tests are considered, and the comparative success that attends their use at present is due more to the homogeneity of human mentality than to the scientific derivation of the measuring instruments.

III. *The Coefficients of Variability of Individual Tests*—The influence of different numbers of units in the various individual tests may be eliminated by reducing the values, which have been the basis for the discussion in the two preceding sections, to coefficients of variability in the same way that we calculated the coefficients of variability for differences in total scores. There is one point which must be considered, however, and that is that the probable error of the difference between two averages is affected in a constant direction by the number of cases involved. Although the differences between any two averages probably will not vary markedly with double the number of cases, the probable error of that difference will be smaller, if

twice the number of cases are studied. Where varying numbers of pupils have been tested by the scales, this fact introduces a certain element of unreliability into any comparisons which may be made between the different scales. This point applies to total scores as well as to the scores in the individual tests, although it was not mentioned in the section discussing total scores. Within the same scale, however, this fact ceases to operate because the same number of cases is involved in all the tests of a scale.

The coefficients of variability for the individual tests of the six scales studied are presented in Tables XX to XXV. The wide variations between the individual tests which were noted briefly in our discussion of the differences between successive grade averages have disappeared. The result of reducing them all to the same basis without reference to their effect on the total score shows them to be more nearly alike in discriminative power if the differences in weighting were equalized. Nevertheless, it can still be shown that some of the scales are much better organized than others. In the Otis Scale, for example, Test I shows a coefficient of variability for grades VI to IX (0.78) that is ten times as large as that shown by Test VI (0.076). An extreme range, such as this, is plainly not to the advantage of a scale. The variations in discriminative power between tests should not be considered as evidence that the tests are of no value. They merely mean that in the present form the tests are not well organized for the work that they are to do. It is altogether possible that a revision might introduce the changes that are needed.

TABLE XX—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR THE OTIS SCALE

Test	Coefficients of Variability of Differences between Scores for Indicated Grades								
	6&7	7&8	8&9	9&10	10&11	11&12	6&12	6&9	9&12
I	7.0	0.315	0.285	0.50	0.50	3.0	0.055	0.78	0.37
II	0.50	0.50	0.235	0.666	0.363	0.571	0.074	0.099	0.17
III	0.80	0.296	4.0	0.40	0.125	0.357	0.060	0.127	0.104
IV	0.379	0.478	1.666	0.166	2.50	3.0	0.058	0.078	0.19
V	0.277	5.0	0.187	0.50	0.375	0.363	0.153	0.096	0.397
VI	0.250	0.454	0.50	0.30	1.0	0.30	0.025	0.076	1.10
VII	1.20	0.137	0.60	0.750	1.50	2.0	0.116	0.086	0.854
VIII	5.0	5.0	0.272	0.375	0.33[illegible]	1.50	0.333	0.322	3.37
IX	1.0	0.272	1.750	0.857	0.60	0.714	0.033	0.123	0.224
X	0.625	0.312	0.666	3.0	3.0	[a]	0.20	0.213	1.85

[a] The difference was zero. This made it impossible to compute a coefficient of variability that would mean anything to any but a mathematician.

TABLE XXI—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR CLASSIFICATION TEST

Test	Coefficients of Variability of Differences between Scores for Indicated Grades									
	5&6	6&7	7&8	8&9	9&10	10&11	11&12	5&12	6&9	9&12
I	2.0	0.350	0.692	1.0	0.60	0.428	0.50	0.056	0.062	0.016
II	0.108	1.40	0.080	0.380	0.75	0.333	4.33	0.066	0.098	0.022
III	0.166	0.187	0.571	0.285	0.294	1.0	1.0	0.045	0.010	0.019
IV	0.166	0.727	0.133	0.222	[a]	2.0	0.40	0.055	0.085	3.0
V	0.153	0.111	0.333	1.333	0.411	0.350	1.750	0.061	0.102	0.181
VI	0.184	0.233	0.70	0.152	0.916	0.478	5.50	0.072	0.106	0.333
VII	0.081	0.375	0.444	0.263	1.0	1.50	0.50	0.053	0.022	0.250
VIII	0.125	0.137	2.50	0.104	8.0	0.375	0.555	0.038	0.063	0.016

[a] The difference was zero.

TABLE XXII—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR VIRGINIA DELTA I

Test	Coefficients of Variability of Differences between Scores for Indicated Grades					
	3 & 4	4 & 5	5 & 6	6 & 7	7 & 8	3 & 8
I	0.111	0.266	0.129	0.111	0.190	0.033
II	0.055	0.111	0.055	0.055	[a]	0.015
III	0.333	0.166	0.20	0.285	1.0	0.181
IV	0.069	0.142	0.133	0.181	0.166	0.027
V	0.052	0.047	0.142	0.068	1.0	0.012
VI	0.048	0.105	0.062	0.073	0.214	0.018

[a] The difference was zero.

TABLE XXIII—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR WHIPPLE'S GROUP TEST

Test	Coefficients of Variability of Differences between Scores for Indicated Grades			
	4 & 5	5 & 6	6 & 7	4 & 7
I	0.116	0.113	0.285	0.055
II	0.50	0.166	0.50	0.166
III	0.50	0.227	0.206	0.114
IV-Pt. 1	0.212	0.097	4.0	0.082
Pt. 2	0.714	0.138	0.156	0.071
Pt. 3	0.454	0.267	0.30	0.125
VI [a]	1.0	0.250	0.363	0.115

[a] Test V was not used.

TABLE XXIV—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR PRIMER SCALE

Test	Coefficients of Variability of Differences between Scores for Indicated Grades		
	1 & 2	2 & 3	1 & 3
I	0.103	0.150	0.081
II	0.125	0.136	0.065
III	0.111	0.30	0.086
IV	0.081	0.115	0.047

TABLE XXV—COEFFICIENTS OF VARIABILITY FOR SERIES G AND H BY GRADES FOR VOCABULARY SCALE

Grades	Series G	Series H
3 & 4	0.072	0.089
4 & 5	0.083	0.250
5 & 6	0.096	0.111
6 & 7	0.166	0.133
7 & 8	0.136	0.122
8 & 9	0.272	0.208
9 & 10	0.235	0.161
10 & 11	0.555	0.555
11 & 12	2.000	0.333
3 & 12	0.019	0.019
3 & 6	0.033	0.039
6 & 9	0.041	0.031
9 & 12	0.083	0.075

An examination of these tables reveals the reason for the lack of discriminating power in the high school which is shown by the Otis, Classification, and Vocabulary scales. The coefficients of variability are large. Since these coefficients arise from dividing the probable errors of the differences between grade scores by these differences, large coefficients indicate small differences and large probable errors of these differences. In other words, they indicate little discrimination between the performance of the grades in question and the discrimination that exists is unreliable. When, therefore, relatively large coefficients are found in connection with the high-school grades, it means that the materials are not so graded that they will provide sufficient steps of difficulty for these grades. This suggests, then, that differentiation in the content of intelligence scales will be needed for different grade levels. We shall prob-

ably need as much flexibility in intelligence scales as has been found necessary in educational tests.

Many details are shown by the tables. For example, it will be noted that in Table XXI Test IV in the Classification Test does not appear to be properly graded for grades IX to XII. It may be added parenthetically that Dr. Theisen has been aware of this fact for some time, for the same point was revealed by his own distributions of the scores in this test. Thus, the tables furnish an opportunity for many comparisons. It is not felt that it is worth while to call attention to all of them. The chief irregularities in the tests have been pointed out in the preceding comments.

IV. *Correlations between Equivalent Tests*—Every person who has had even a limited experience with examinations and mental test work realizes that individuals do not always do their best when placed in a test situation. This variability has been noted many times in the literature of school marks and similar studies. But since formal tests have become rather widely used, many people have neglected this fact and have assumed that for all intents and purposes the scores made by individuals on a test are reliable to a high degree. This assumption is not always true.

This investigation afforded an opportunity for a consideration of this question. Several of the scales contained tests that could be considered equivalent. In some cases there was a divergence in form and structure of the test which might be responsible for variability, but the resemblance was close enough to make it seem fair to make the comparison. Accordingly, the correlations presented in Table XXVI were computed.

TABLE XXVI—CORRELATION BETWEEN EQUIVALENT TESTS

Test	Grade					
	III	IV	V	VI	VII	VIII
Arithmetic in Otis and Classification						0.72±0.05
Otis and Virginia				0.52±0.06	0.53±0.12	0.51±0.07
Classification and Virginia			0.58±0.05	0.35±0.07	0.43±0.07	0.70±0.05
Trabue in Classification and Group			0.36±0.06			
Analogies in Classification and Otis						0.11±0.10
Opposites in Otis and Classification						0.36±0.11
Otis and Virginia				0.19±0.11	0.56±0.11	0.28±0.08
Classification and Virginia			0.45±0.06	0.34±0.07	0.47±0.06	0.46±0.06
Vocabulary—Series G & H	0.41±0.05	0.20±0.04	0.38±0.04	0.40±0.04	0.58±0.04	0.55±0.04
Information in Virginia and Classification			0.58±0.05	0.54±0.05	0.62±0.06	0.39±0.07
Common Sense in Virginia and Classification			0.51±0.06	0.30±0.06	0.39±0.08	0.14±0.08

A casual inspection of this table shows that there is a great difference in the degree of reliability between comparable individual tests. Unless the repetition of a test produces approximately the same score for individuals when taken on successive occasions, its reliability may be seriously questioned. When individual tests do not show a fairly high degree of reliability as measured by the correlation between the individual scores made on the two trials, one cannot expect a scale composed of these unreliable individual tests to be very satisfactory.

It is possible that tests differing in structure and content (e.g., Arithmetic and Analogies) may measure the same mental functions. The question then arises as to whether it is better to give different tests or merely to use different forms of the same test. If the same test is repeated without rest periods, the results may not be as satisfactory as if different tests are used, because children may not concentrate as readily on the later repetitions of the test. On the other hand, later repetitions of a test might secure better results, if novelty was a distracting factor during the first attempt. For variety's sake—and variety is the spice of schoolroom testing—it may be desirable to use tests different in character. In all probability, if the scale is to be repeated, the best plan will be to administer the different forms on successive days. By this procedure the chance physiological variations which no doubt affect the child's efficiency will tend to be neutralized; for the average of several performances measured on different days usually gives more accurate data about a pupil than a single measure. This point is one that lends itself to further study and is one that ought to be made the center of careful research in order that we may have an adequate foundation for both intelligence and educational testing.

This point may be considered from another angle. Brown [6] has given us a formula by which we can determine the number of repetitions of equivalent tests needed to secure a reliability of any desired value. With this formula the number

[6] Brown, Wm. *Mental Measurement*, Cambridge University Press, 1911, pp. 101-102. Brown's formula for computing the reliability of the amalgamation of the scores of similar tests is

$$r_n = \frac{nr_1}{1+(n-1)r_1}; \text{ whence } n = \frac{r_n(r_1-1)}{r_1(r_n-1)}$$

of repetitions of a test needed to secure a reliability of 0.80, 0.90, and 0.95 have been computed and are presented in Table XXVII. From this table we may readily conclude that equivalent tests which do not reveal a correlation with one another of at least 0.50 are of doubtful value. Applying the values in this table to the correlation coefficients presented for equivalent tests in Table XXVI, it can be seen that a large number of the individual tests are of poor reliability if they are to be used to decide the

TABLE XXVII. NUMBER OF REPETITIONS OF A TEST NEEDED TO SECURE RELIABILITY INDICATED [a]

Correlation Coefficient	Reliability		
	0.80	0.90	0.95
0.30	10	21	45
0.35	8	17	36
0.40	6	14	29
0.45	5	11	24
0.50	4	9	19
0.55	4	8	16
0.60	3	6	13
0.65	3	5	11
0.70	2	4	9
0.75	2	3	7
0.80		3	5
0.85		2	4
0.90			3

[a] In some cases the number of repetitions given will result in a reliability slightly better than that indicated. Round numbers alone are presented because it was felt to be absurd to speak of fractional repetitions.

fate of individuals. When tests are used to measure groups they are more reliable. This conclusion is supported by the only correlation coefficient which could be computed from these data—a value of 0.74 ± 0.05, representing the correlation between the room scores in the elementary grades for Series G and Series H, Sentence Vocabulary Scale.

V. *Zero Scores in Individual Tests*—Zero scores result from tests which are not suited to the capacities of the child. Whenever such scores become appreciable they imply that the test is being administered at a point beneath its lower level of efficiency. The percents of zero scores for the individual tests in the different scales are presented in Table XXVIII. Attention is called to those which have especially high percents of such scores. They are tests IV and IX, Otis Scale; tests III and VII, Classification Test; tests IV and V, Virginia Delta I; and

TABLE XXVIII. PERCENT OF ZERO SCORES IN INDIVIDUAL TESTS FOR DIFFERENT SCALES

Scale or Test	Percent of Zero Scores in Indicated Grades							
	I	II	III	IV	V	VI	VII	VIII
OTIS								
I Following Directions	.....	.....	.....	.....	.....	0	0	0
II Opposites	.....	.....	.....	.....	.....	0	0	0
III Disarranged Sentences	.....	.....	.....	.....	.....	2	0	0
IV Proverbs	.....	.....	.....	.....	.....	8	11	11
V Arithmetic	.....	.....	.....	.....	.....	0	0	0
VI Geometric Figures	.....	.....	.....	.....	.....	0	0	0
VII Analogies	.....	.....	.....	.....	.....	0	0	0
VIII Similarities	.....	.....	.....	.....	.....	0	0	0
IX Narrative Completion	.....	.....	.....	.....	.....	8	11	0
X Memory	.....	.....	.....	.....	.....	0	0	0
CLASSIFICATION								
I Following Directions	.....	.....	.....	.....	0	0	0	0
II Synonym-Antonym	.....	.....	.....	.....	0	0	0	0
III Arithmetic	.....	.....	.....	.....	8	0	0	0
IV Common Sense	.....	.....	.....	.....	1	0	0	0
V Completion	.....	.....	.....	.....	4	0	0	0
VI Analogies	.....	.....	.....	.....	1	0	0	0
VII Number Completion	.....	.....	.....	.....	40	8	13	17
VIII Information	.....	.....	.....	.....	1	0	0	0
VIRGINIA DELTA I								
I True-False	.....	.....	1	0	0	0	0	0
II Arithmetic	.....	.....	6	0	0	0	0	0
III Picture Completion	.....	.....	0	0	0	0	0	0
IV Synonym-Antonym	.....	.....	25	18	6	55	2	0
V Common Sense	.....	.....	17	5	1	1	0	0
VI Information	.....	.....	8	0	0	0	0	0
WHIPPLE								
I Arithmetic	.....	.....	.....	0	0	0	0	
II Completion	.....	.....	.....	0	0	0	0	
III Substitution	.....	.....	.....	4	6	0	3	
IV Reasoning Tests—								
Part I	.....	.....	.....	0	0	0	0	
Part II	.....	.....	.....	65	48	30	3	
Part III	.....	.....	.....	40	30	18	9	
V Punched Hole Test	.....	.....	.....					
VI Proverbs	.....	.....	.....	16	23	6	0	
PRIMER								
I Dot Pattern	3	0	2					
II Classification	1	1	1					
III Form Board	1	2	0					
IV Absurdities	3	1	0					
VOCABULARY								
Series G	.....	.....	10	0	0			
" H	.....	.....	14	1.5	0			

Test IV, Parts II and III, Whipple's Group Test. There is no apparent agreement in respect to zero scores among tests of the same kind occurring in different scales. Those which do not qualify in some scales apparently do so in others. The difference in efficiency of these various tests can be attributed to their degree of difficulty and the way they are administered. The general information presented by the Vocabulary and Virginia Delta I scales seems to suggest that literary scales have a low degree of efficiency in the third grade. There are many children in this grade who have not sufficiently mastered the mechanics of literary tests to permit the measurement of their intelligence by these instruments. A zero score on a test does not imply zero mentality. It does imply an inferior mentality or lack of comprehension. In general, it may be said that zero scores when they are appreciably numerous indicate the approximate lower limits of the efficiency of a test. If a scale contains several tests which result in zero scores in the same grade, it is clear that the resulting total score cannot justly be compared with total scores wherein all of the tests function. What effect this point would have on the discriminative power of a scale is not answered by these data. It raises a question which might well be given careful study.

The question of zero scores should also be considered from the administrative point of view. Whenever a test is given which the child does not comprehend because of incomplete instruction or difficult subject matter, the problem of copying is introduced. Nearly every pupil is accustomed to take cues from others when anything is not clear. Consequently, when children in groups are asked to do something which is beyond them, they look around to see what others are doing. This behavior is not such dishonesty as would warrant a severe penalty. However, it is an argument against the administration of tests which result in many zero scores.

F—General Comparison of Scales

The foregoing data in Part II furnish many items of information concerning the six intelligence scales considered. Isolated, these facts do not tell us very much; but when summarized they prove more valuable. It will be worth our while to spend a few moments discussing the criteria by which scales may be considered.

I. *Grades for which the Scales are Suitable*—If some of the tests of a scale result in zero scores, it is not advisable to administer that scale in grades below the one in which zero scores first appear in quantity because the percent of zero scores increases very rapidly in successively lower grades and because administrative difficulties are raised when pupils are not kept busy by a test. The upper limit of a test is determined by the difficulty of the material. If total scores do not discriminate between grades, the upper limit has been reached.

II. *Ease of Administration*—Scales which have complicated directions, requiring careful study and much practice on the part of the examiner before they can be administered with any guarantee of success, are not so desirable as scales which are provided with more simple administrative procedures. Scales requiring a long time are preferable only when they are more discriminative than those requiring less time.

III. *Difficulties Involved in Scoring*—Checking the work performed by the pupils is a purely clerical task which is much the same for the different scales. Sufficient information will be furnished on this point by a mere comparison of the amounts of time required to score an individual paper.

IV. *Correlation with Scholarship*—As has already been pointed out, these correlations are rather low. Nevertheless, it is felt that the scale which correlates the highest is the best, and vice versa.

V. *Reliability*—Scales which do not reveal sufficiently large differences between the average scores of successive grades manifestly cannot discriminate between the different abilities presented by the children in those grades. When differences are accompanied by large probable errors the reliability of such differences is low—i.e. the discrimination, whatever it may be, is questionable.

VI. *Individual Tests*—Finally, the scales may be considered from the standpoint of the degree of balance shown by the different individual tests which compose them. If the individual tests vary markedly in their contributions to the total score and show inversions between averages for successive grades, the scales in which such tests occur are not as valuable as those which are well balanced and regular.

Using the criteria which we have set up in the preceding

TABLE XXIX—GENERAL COMPARISONS OF THE SIX SCALES

Scale	Range of Suitability (Grades)	Administration	Scoring	Correlation with Scholarship	Reliability of Total Scores between Successive Grades	Individual Tests
Otis	6—12	Slightly complicated and lengthy	Lengthy	0.42	poor	Erratic and poorly-balanced
Classification	5—12	Slightly complicated and of medium length	Medium length	0.50	poor	Regular but poorly-balanced
Virginia	3— 8	Simple and brief	Brief	0.59	good	Regular but poorly-balanced
Group	6— 8	Simple but lengthy	Very lengthy	[a]	[a]	Regular and well-balanced
Primer	1— 3	Simple and brief	Very brief	0.34	fair	Regular and well-balanced
Vocabulary	3—12	Very simple and brief	Brief	0.47	good	Regular and well-balanced

[a] Data not available.

paragraphs, we may compare the six scales as indicated in Table XXIX. It is not always easy to generalize, and this limitation must be kept in mind as the comparisons are examined. To some extent they are personal opinions.

G—Summary and Comments

I. The present intelligence scales admit of much improvement. Some of them are erratic and poorly balanced, with comparatively poor reliability for the total scores. Others require too much time for administration or scoring. Nearly all of them can be improved by the addition of new material or by the preparation of different tests for different intelligence levels.

II. The best scales for intermediate and grammar grades seem to be the Virginia Delta I and the Vocabulary scales. A combination of these two seems to offer the best measuring instrument.

III. The Classification Test seems best for the high-school grades.

IV. The Primer Scale is well organized from the standpoint of administration, scoring, and balance, but it is of questionable diagnostic value. It should always be supplemented by other tests.

V. The value of the different intelligence scales should not be determined alone on statistical grounds. There are so many factors that influence the work of the child that it is impossible in every instance to forecast his performance in school work accurately by means of intelligence tests. There will be exceptions in most classes. These exceptions should be analyzed in the light of the information available and, in this way, many apparent deviations between scholarship and intelligence ratings may be satisfactorily explained.

VI. The present outlook for the derivation and use of group intelligence scales is good. The field is fertile and there is every indication that successful scales may be prepared.

PART III—MENTAL SURVEY OF THE CHAMPAIGN PUBLIC SCHOOLS

A—Introduction

The point of view taken in this portion of this bulletin is quite different from that assumed in Part II where our attention was directed constantly to the individual scales and tests. In Part III, on the contrary, we shall consider the child as the unit. Intelligence scales function only as they throw light on the mental ability of children.

For the purposes that we have in mind in Part III it will be most convenient to make the assumption that for practical purposes the tests which have been used are fairly satisfactory. Of course, in the light of what has been pointed out in Part II, an assumption of this sort with respect to some of the test scores would not be warranted if it were necessary to make individual recommendations on the basis of these values. This procedure, however, will give us a basis of treatment which may be copied by superintendents and teachers who wish to analyze the intelligence situations in their schools. With this point of view in mind, the statistical treatment in this section will be very simple. Everything will be determined on the basis of distributions, with the median as the measure of central tendency.

B—Variations between Schools and Grades

Teachers often remark that the pupils whom they teach in one semester are not equal in mental ability to those whom they taught in some particular preceding semester. Sometimes educators are inclined to charge these opinions on the part of the teachers to personal idiosyncracies, assuming that all school grades are much alike in character. This assumption is seldom justified. Classes vary from year to year in their composition. One may have a large number of mediocre pupils in it, with comparatively few bright or dull ones. Another may have the extremes with few average pupils. In other words, classes vary considerably when the finer points are considered. On the other hand, teachers often erroneously assume that there is an intel-

ligence difference between two classes. They are often led to make this assumption by the responses of a few superior or inferior pupils. A few exceptionally bright or dull pupils in a room will influence the tone of a group in a manner out of all proportion to their number.

I. *Variations as Shown by Median Scores in Intelligence Scales*—The variation between grades in the different schools may be shown by the differences between median scores made by the children of those grades in the different intelligence scales. These median scores are presented in Tables XXX and XXXI.

An examination of Tables XXX and XXXI shows that the same grades in the different schools are much alike in general intelligence when considered from the standpoint of the class medians. The median scores show some variation, a part of which may be due to their unreliability as measures of the group intelligence, and a part of which may be due to real differences in the grades tested. In general, however, there is a surprising uniformity and the differences which appear consistently are the ones which have been recognized by the teachers and supervisors. School No. 2 is in the best section of the city of Champaign, Illinois, and School No. 5 draws partly from the poorest section of the city. The differences between the median scores for the grades of these two schools show a decided superiority in both the Vocabulary and Virginia scales in favor of School No. 2 in every grade. If we examine School No. 9 (Table XXXI), we find that the differences between rooms are very noticeable. The pupils in the six rooms in the eighth grade had been classified into sections according to their scholarship records. At the beginning of the school year their records in the seventh grade had been used as a rough basis for classification. When the year had progressed far enough to give examinations, the pupils had been tested very carefully, and a reclassification had been made on the basis of these examination records. As a consequence, Room No. 4 contained mainly the superior pupils, while Room No. 3 contained the poorest section. The classification of the pupils on the basis of scholarship resulted in some overlapping between the different sections. The teachers recognized this fact, since they rated some of the children in Room No. 3 as of average or above average ability in scholarship. The ranks of the different sections based on scores in the Vocabulary and the Virginia scales agree closely with the ranks previously given these sections by the school officials.

TABLE XXX—SCHOOL SCORES FOR EACH SCALE BY GRADES

School and Scale	Grade						
	I	II	III	IV	V	VI	VII
School No. 1							
Vocabulary	.....	.....	22	29	38	42	47
Virginia	.....	.....	59	72	86	96	111
Primer	50	62	72				
School No. 2							
Vocabulary	.....	.....	20	33	31	43	48
Virginia	.....	.....	.....	74	89	105	115
Primer	48	53	56				
Classification	.....	.....	.....	.....	59	83	97
School No. 3							
Vocabulary	.....	.....	.....	30	36	38	47
Virginia	.....	.....	.....	73	79	100	118
Primer	40	54	65				
Otis	.....	.....	.....	.....	.....	94	117
School No. 4							
Vocabulary	.....	.....	14	32	34	42	
Virginia	.....	.....	44	63	81	92	
Primer	43	53	60				
School No. 5							
Vocabulary	.....	.....	.....	28	30	38	44
Virginia	.....	.....	40	72	74	88	105
Primer	41	54	62				
Classification	.....	.....	.....	.....	58	81	100
School No. 6							
Vocabulary	.....	.....	8	30	31	30	45
Virginia	.....	.....	.....	65	79	89	107
Primer	46	54	66				
School No. 7							
Vocabulary	.....	.....	17	28			
Virginia	.....	.....	48	58			
Primer	37	55	53				
School No. 8							
Vocabulary	.....	.....	17	24			
Virginia	.....	.....	43	66			
Primer	41	62	70				

II. *Quartile Variation of Children between Schools by Grades*—The preceding section, which considered merely the median scores for the different rooms, gave a general picture of the situation in the city. A general picture, however, is not entirely adequate. Two fourth-grade classes may have identical median scores but, when the individual children are considered, there may be important differences. One may have children all of whom are of approximately the same mental ability, the

TABLE XXXI—SCORES IN SCHOOL NO. 9 FOR EACH SCALE BY ROOMS

Grade and Scale	Room					
	1	2	3	4	5	6
Seventh Grade						
Vocabulary	45	43				
Virginia	111	106				
Eighth Grade						
Vocabulary	56	55	43	63	51	46
Virginia	123	120	104	136	113	119
Classification	.....	.....	.....	116	103	
Otis	.....	.....	.....	143		121

other may have a number of dull children balanced by a corresponding number of bright children. Manifestly, it will be much easier to teach the first class than the second. No statistical presentation can take the place of the detail furnished by the distribution table. It can be used, however, to best advantage only when a few groups of children are under observation. Under the present circumstances some general expression of variation is needed—something that will give an average "scatter" just as the median gives an average score. We shall use the quartile deviation for this purpose.

First we may regard the pupils of each grade as constituting a single group. Table XXXII shows the median and quartile deviation (half the scale distance between the 25- and 75-percentile) for each test and for each grade.

Second, we may consider the deviation of each pupil from the median of his grade. Clearly if we wish to handle these deviations together we must express them in a common or at least a comparable unit. We propose to use the quartile deviations given in Table XXXII as such units. If, for example, a fourth-grade pupil scored 43 in the Vocabulary Scale, his deviation from the fourth-grade median for that test (31) would be +12. Since the corresponding quartile deviation is 6, his deviation in terms of the quartile deviation would be +2. Similarly if a sixth-grade pupil scored 76 in the Virginia Scale, his deviation in terms of the units of the scale would be −22. One might think that this deviation is greater than that of the fourth-grade pupil just mentioned. It is true that numerically and in terms of scale units it is nearly twice as great. But the variability of sixth-grade scores on the Virginia Scale is in general

greater, amounting, according to Table XXXII, to 11. In other words, the expectancy of deviation from the median is greater. This makes the larger numerical deviation of 22 of less significance. As before we may express the relation between the individual deviation and the general measure of dispersion for the group by dividing the former by the latter. When this is done we find that, in terms of the quartile deviation for the grade and test, the individual deviation for the sixth-grade pupil whom we are considering would be −2.

TABLE XXXII. MEDIANS AND QUARTILE DEVIATIONS

Scale	Grade							
	I	II	III	IV	V	VI	VII	VIII
Vocabulary								
Median	.....	.....	21	31	33	40	47	53
Quartile Deviation	.....	.....	7	6	6	6	7	7
Virginia								
Median	.....	.....	48	68	80	98	110	117
Quartile Deviation	.....	.....	10	10	11	11	12	11
Classification								
Median	.....	.....	.....	.....	59	83	99	110
Quartile Deviation	.....	.....	.....	.....	12	10	12	11
Otis								
Median	.....	.....	.....	.....	.....	94	117	128
Quartile Deviation	.....	.....	.....	.....	.....	10	16	13
Picture Completion								
Median	19	24	26	28				
Quartile Deviation	4	4	3	2				
Primer								
Median	44	57	64					
Quartile Deviation	9	7	7					

Table XXXIII gives the distribution of pupils of each grade according to the deviations of their scores from the median, deviations being expressed in terms of the quartile deviation for the grade to which the pupil belonged and for the test which he took. If the same pupil took more than one test the average of his deviations is given. Thus each child is entered but once. The medians shown in Table XXXIII are the medians of deviations.

These medians show in a general way the differences between the performances of pupils in the several grades. The differences in the composition of each grade of the various

TABLE XXXIII—DISTRIBUTION OF PUPILS IN EACH GRADE AND SCHOOL ACCORDING TO THEIR DISPERSION FROM THE MEDIAN. THE UNIT IS THE QUARTILE DEVIATION

Grade	Deviation $(\frac{x}{Q})$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
I									
	+3.0 to +3.9	1							
	+2.0 to +2.9	3			3				
	+1.0 to +1.9	3	2	3	7	2	2	1	2
	0 to +0.9	18	10	5	15	5	9	4	12
	−1.0 to −0.1	6	2	16	9	13	5	1	10
	−2.0 to −1.1	3	1	7	8	5	1	1	5
	−3.0 to −2.1		4	2	1	4		2	1
	−4.0 to −3.1					1		1	
	Total	34	19	33	43	30	17	10	30
	Median	+0.44	+0.25	−0.53	+0.25	−0.62	+0.28	0	−0.10
II									
	+3.0 to +3.9								
	+2.0 to +2.9		1						
	+1.0 to +1.9	11	4	2	1	3	2	4	4
	0 to +0.9	16	15	13	13	7	9	6	6
	−1.0 to −0.1	7	13	15	15	10	0	2	9
	−2.0 to −1.1	1	2	1	1	2	3	2	7
	−3.0 to −2.1	2	1	2		1	3	1	2
	−4.0 to −3.1			2		2	1	4	
	−5.0 to −4.1							1	
	Total	37	36	35	30	25	18	20	28
	Median	+0.53	+0.13	−0.17	−0.07	−0.25	+0.22	0	−0.44
III									
	+3.0 to +3.9	1							
	+2.0 to +2.9	3	1	1					
	+1.0 to +1.9	7	3	5	1	2	1	1	2
	0.0 to +0.9	14	11	19	8	12	6	7	12
	−1.0 to −0.1	9	10		14	7	3	6	11

TABLE XXXIII—(*Continued*)

Grade	Deviation $\left(\frac{x}{Q}\right)$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
	−2.0 to −1.1	2	5	10	6	6	2	1	4
	−3.0 to −2.1		2	1	4	1	3	3	
	−4.0 to −3.1		2		1	2		1	
	−5.0 to −4.1							1	
	Total	36	34	36	34	30	15	20	29
	Median	+0.50	−0.20	+0.37	−0.57	−0.17	−0.14	−0.33	−0.05
IV									
	+4.0 to +4.9	1	3						
	+3.0 to +3.9	0	1						
	+2.0 to +2.9	2	3		1				1
	+1.0 to +1.9	3	5	5	3	1	2		0
	0 to +0.9	12	9	12	16	14	2		2
	−1.0 to −0.1	16	8	6	7	8	6		9
	−2.0 to −1.1	0	4	6	7	1	2		2
	−3.0 to −2.1	2	4	7		1	1		2
	−4.0 to −3.1					1			1
	Total	36	37	36	34	26	13		17
	Median	0	+0.28	−0.17	+0.16	+0.14	−0.42		−0.61
V									
	+3.0 to +3.9	1	1				1		
	+2.0 to +2.9	3	1	3	2	2	1		
	+1.0 to +1.9	7	10	2	4	5	1		
	0 to +0.9	7	9	13	23	12	9		
	−1.0 to −0.1	12	8	9	17	24	6		
	−2.0 to −1.1	1	3	3	9	8	3		
	−3.0 to −2.1		3			4	3		
	−4.0 to −3.1		1				1		
	Total	31	36	30	55	55	25		
	Median	+0.36	+0.33	+0.23	+0.07	−0.35	−0.08		

TABLE XXXIII—(*Continued*)

Grade	Deviation $\left(\frac{x}{Q}\right)$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
VI									
	+3.0 to +3.9		1	2					
	+2.0 to +2.9	3	6	4		1			
	+1.0 to +1.9	3	4	4	5	4	1		
	0 to +0.9	13	14	16	20	7	2		
	−1.0 to −0.1	5	8	16	19	13	6		
	−2.0 to −1.1	5	4	6	12	8	6		
	−3.0 to −2.1	2	1	6	1	0	0		
	−4.0 to −3.1	1				1	1		
	Total	32	38	54	57	34	16		
	Median	+0.23	+0.43	−0.06	−0.18	−0.38	−0.83		
VII									
	+2.0 to +2.9	4	4	2					
	+1.0 to +1.9	2	5	2		3	3		
	0 to +0.9	8	13	6		10	4		
	−1.0 to −0.1	12	11	6		13	3		
	−2.0 to −1.1	4	4	3		4	4		
	−3.0 to −2.1	1	1	1		3	2		
	−4.0 to −3.1	1							
	Total	32	38	20		33	16		
	Median	−0.17	+0.23	0		−0.27	−0.33		

schools is made evident by the distributions. Some schools show a large number of below-average children, while others show the opposite condition. School No. 9, where the eighth grade was divided into sections based on scholarship achievement, shows (Table XXXIV) the decided superiority of Room No. 4 and the marked inferiority of Room No. 3. The overlapping between these rooms indicates the insufficiency of the scholarship basis when used in the classification of children into sections. The quartile deviation provides a convenient device for showing the intelligence composition of rooms and grades.

TABLE XXXIV. DISTRIBUTION OF PUPILS IN SCHOOL NO. 9 BY ROOMS AND ACCORDING TO THEIR DISPERSION FROM THE MEDIAN. THE UNIT IS THE QUARTILE DEVIATION

Grade	Deviation $(\frac{x}{Q})$	Room 1	Room 2	Room 3	Room 4	Room 5	Room 6
VII	+1.0 to +1.9	4	4				
	0 to +0.9	8	7				
	−1.0 to −0.1	9	9				
	−2.0 to −1.1	5	7				
	−3.0 to −2.1	2	1				
	−4.0 to −3.1	1	1				
	Total	29	29				
	Median	−0.28	−0.39				
VIII	+2.0 to +2.9	2	3		4	1	
	+1.0 to +1.9	7	10	3	12	3	3
	0 to +0.9	14	7	1	10	7	9
	−1.0 to −0.1	6	11	5	2	9	8
	−2.0 to −1.1	4	2	16	3	8	11
	−3.0 to −2.1			3		2	2
	Total	33	33	28	31	30	33
	Median	+0.46	+0.50	−1.31	+1.04	−0.44	−0.56

III. *Classification of Children on the Basis of the Intelligence Quotient.*

a. The quartile deviation of children takes into account only their present intelligence status—It is almost a common opinion that two people may be rated equal in intelligence, although they may be widely different in their ability to profit by their future experience. We may make this clear by a concrete illustration. Clarence S. and Frank V. are in the fourth grade and both secure the same scores when given intelligence tests.

Clarence is 16 years old chronologically, while Frank is only 9. Manifestly, these two pupils have vastly different educational prospects. The older boy probably has reached his intellectual maturity and will soon drop out of school. He is a retarded pupil with an intelligence quotient of approximately 0.75. On the other hand, the younger child is of superior ability with an intelligence quotient of 1.10. His future educational prospects are bright. He will complete the elementary school; and if he enters high school and college, he will probably succeed. Clarence is the freight train that has started years before, while Frank is the express that has overtaken him in his educational journey.

Before it is possible to convert the scores in the intelligence tests into mental age values, it becomes necessary to establish standards for the different scales. In doing this it is necessary to make the commonly used assumption that the average individual reaches his intellectual maturity at the age of sixteen. Further, it is assumed that the highest intellectual development is represented by nineteen years' mentality. With these hypothetical bases, it is a comparatively easy matter to set up age standards for the different scales. Although the data available were incomplete at the upper and lower ranges, it is felt that an adequate allowance was made for the selected nature of the groups from which the standards were obtained. The standards used in determining the intelligence quotients of the children are presented in Table XXXV for the different scales. Below the score for each age group are two figures in italics. With these figures it is possible to interpolate ages in years and months. These standards may be used as follows: If a child makes a score of 42 in the Vocabulary Scale, he has an approximate mental age of 13.3 years. If his score is 100 in the Virginia Scale, his mental age is 13.4 years.

b. Distributions of children by intelligence quotients— The standards presented in Table XXXV made it possible to determine the mental age of each child as revealed by the scores that he made in the tests. Here again it was felt that the average of several figures is a better index than any one of them; hence the mental ages given by the various test scores were averaged for each of the children. This average mental age was divided by the chronological age, giving the intelligence quotient. In the high school where the children were older than 16 years chronologically, it was necessary to use 16 years as the divisor, since

TABLE XXXV. STANDARD SCORES FOR INDICATED AGES

Scale	4 yrs. to 4.9 yrs.	5.0–	6.0–	7.0–	8.0–	9.0–	10.0–	11.0–	12.0–	13.0–	14.0–	15.0–	16.0–	17.0–	18.0–	19.0–
Vocabulary					5 *0-10*	15 *11-18*	22 *19-25*	29 *26-32*	36 *33-39*	43 *40-46*	50 *47-53*	57 *54-60*	65 *61-69*	74 *70-78*	82 *79-86*	90 *87-93*
Virginia				15 *0-25*	36 *26-44*	53 *45-59*	65 *60-70*	76 *71-81*	87 *82-93*	99 *94-107*	115 *108-123*	129 *124-137*	145 *138-155*			
Classification						15 *0-25*	35 *26-42*	50 *43-57*	65 *58-71*	78 *72-85*	93 *86-99*	106 *100-114*	122 *115-132*	140 *133-152*	163 *153-174*	186 *175-197*
Otis							40 *20-52*	65 *53-74*	85 *75-93*	102 *94-108*	118 *109-126*	135 *127-141*	148 *142-155*	162 *156-169*	176 *170-183*	190 *184-195*
Primer	6 *0-12*	20 *13-26*	32 *27-37*	43 *38-48*	54 *49-60*	66 *61-72*	78 *73-83*	86 *84-89*								

TABLE XXXVI. DISTRIBUTION OF INTELLIGENCE QUOTIENTS BY SCHOOLS AND GRADES

Grade	I. Q.	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
I	0.50–0.59							3		3
	0.60–0.69					2		0	1	3
	0.70–0.79			1		2	1	2	0	6
	0.80–0.89		3	2		4	0	0	0	9
	0.90–0.99		2	5		2	1	0	3	13
	1.00–1.09		4	7		5	6	1	4	27
	1.10–1.19		8	3		6	2	3	2	24
	1.20–1.29		5	1		3	5	1	1	16
	1.30–1.39		2	3			1	1	1	8
	1.40–1.49		1							1
	Total		25	22		24	16	11	12	110
	Median		1.14	1.04		1.04	1.10	1.05	1.05	1.08
II	0.50–0.59					2				2
	0.60–0.69			1		1		1		3
	0.70–0.79		1	0		1	1	1		4
	0.80–0.89		0	2	3	1	3	2		11
	0.90–0.99		3	9	4	2	1	1		20
	1.00–1.09		5	10	10	4	2	2		33
	1.10–1.19		12	6	3	3	5	3		32
	1.20–1.29		5	8	2		3	1		19
	1.30–1.39		1	1				1		3
	1.40–1.49							1		1
	Total		27	37	22	14	15	13		128
	Median		1.14	1.07	1.04	1.00	1.11	1.08		1.07
III	0.60–0.69			1	1	1		1		4
	0.70–0.79	1	1	1	1	1	1	4	1	11
	0.80–0.89	3	1	1	6	2	1	2	5	21
	0.90–0.99	5	6	6	9	6	2	4	7	45
	1.00–1.09	6	8	8	12	9	2	3	7	55
	1.10–1.19	6	10	7	3	5	2	3	7	43
	1.20–1.29	6	4	3	1	2	1	2	1	20
	1.30–1.39	3	2		1	1	1	1		9
	1.40–1.49		2							2
	Total	30	34	27	34	27	10	20	28	210
	Median	1.10	1.11	1.06	1.00	1.04	1.05	0.98	1.01	1.04
IV	0.60–0.69					1		1	1	3
	0.70–0.79	1		2	2	2		2	1	10
	0.80–0.89	2	1	5	5	6		4	3	26
	0.90–0.99	5	4	5	7	6	1	4	7	39
	1.00–1.09	6	8	7	8	5	3	2	5	44
	1.10–1.19	7	8	8	6	5	1	2	2	39
	1.20–1.29	5	8	5	4	1	2	1	1	27
	1.30–1.39	2	2	4	2			1	2	13
	1.40–1.49	2	4	1						7
	1.50–1.59	1	2							3
	Total	31	37	37	34	26	7	17	22	211
	Median	1.12	1.17	1.09	1.04	0.97	1.08	0.93	0.99	1.06

TABLE XXXVI—(*Continued*)

Grade	I. Q.	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
V	0.50–0.59		1	1		1	1			4
	0.60–0.69	1	1	1		3	1			7
	0.70–0.79	2	0	3	3	7	2			17
	0.80–0.89	3	3	5	7	8	3			29
	0.90–0.99	4	5	9	9	7	4			38
	1.00–1.09	7	7	9	13	8	4			48
	1.10–1.19	7	9	5	7	6	5			39
	1.20–1.29	4	7	3	2	2	3			21
	1.30–1.39	2	2	3	2	2	2			13
	1.40–1.49	1	1			2				4
	Total	31	36	39	43	46	25			220
	Median	1.08	1.11	1.01	1.02	0.96	1.04			1.03
VI	0.60–0.69	2		1	1		1			5
	0.70–0.79	3	1	2	6		1			13
	0.80–0.89	4	2	9	8	4	2			29
	0.90–0.99	5	3	13	12	9	4			46
	1.00–1.09	7	8	11	17	9	5			57
	1.10–1.19	7	10	10	10	7	3			47
	1.20–1.29	4	7	5	4	4	1			25
	1.30–1.39	1	5	1	1	1				9
	1.40–1.49		2	1	0					3
	1.50–1.59				1					1
	Total	33	38	53	60	34	17			235
	Median	1.04	1.15	1.01	1.02	1.04	1.01			1.04
VII								School No. 9		
								Room 1	Room 2	
	0.50–0.59							1		1
	0.60–0.69	1					1	0		2
	0.70–0.79	1	1	2		2	2	6	2	16
	0.80–0.89	5	5	1		7	2	6	4	30
	0.90–0.99	7	6	5		9	2	4	9	42
	1.00–1.09	9	9	7		12	4	5	6	52
	1.10–1.19	4	11	2		2	4	5	3	31
	1.20–1.29	5	4	2		1	1	3	4	20
	1.30–1.39	1	2							3
	Total	33	38	19		33	16	30	28	197
	Median	1.03	1.08	1.02		0.98	1.03	0.95	0.99	1.01

it has been assumed that the average person reaches his mental maturity at this age. These intelligence quotients may be distributed to show the composition of the different grades in each school. It might be added that these values are not so reliable for children who scored low as they are for those who scored high. A child may score low through other factors than the lack of intelligence. Consequently, there are somewhat more children

TABLE XXXVII—DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN THE EIGHTH GRADE OF SCHOOL NO. 9 BY ROOMS

I. Q.	Room Number						Total
	1	2	3	4	5	6	
0.70–0.79			2			1	3
0.80–0.89	1	1	7		3	4	16
0.90–0.99	4	3	10	2	7	10	36
1.00–1.09	11	9	6	6	10	10	52
1.10–1.19	11	12	2	12	8	5	50
1.20–1.29	4	5	1	9	3	2	24
1.30–1.39	1	3		2		1	7
Total	32	33	28	31	31	33	188
Median	1.10	1.13	0.95	1.16	1.08	1.01	1.07

TABLE XXXVIII—DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN THE HIGH SCHOOL BY GRADES

I. Q.	Grade				Total
	IX	X	XI	XII	
0.70–0.79				1	1
0.80–0.89	4	3	3	0	10
0.90–0.99	40	41	23	16	120
1.00–1.09	44	40	48	33	165
1.10–1.19	43	37	41	38	159
1.20–1.29	18	8	15	17	58
1.30–1.39	8	4	2	1	15
1.40–1.49	2				2
Total	159	133	132	106	530
Median	1.08	1.06	1.08	1.11	1.08

showing an intelligence quotient equivalent to that of defective mentality than the facts in the case probably warrant. The excess, however, is not thought to be large; because, with the exception of the first two grades, the intelligence quotient is the combined value of the scores from several scales. As a rule, these scores varied but little with the very low-grade children, although the tests were given at different times, usually weeks apart.

The intelligence quotient is extremely significant. Although it has not been definitely established, the opinion seems to be that it remains approximately constant through life. Occasional exceptions seem to appear, but it is probable that the rule holds as steadily as do most rules regarding mental or physical characteristics. On the basis of the size of the intelli-

gence quotient a number of classifications have been made. Terman's is commonly accepted. It is given in Table XXXIX.

TABLE XXXIX—IMPLICATIONS OF INTELLIGENCE QUOTIENTS

I. Q.	Classification
Above 140	"Near" genius or genius
120 to 140	Very superior intelligence
110 to 120	Superior Intelligence
90 to 110	Normal or average intelligence
80 to 90	Dullness, rarely classifiable as feeblemindedness.
70 to 80	Borderline deficiency, sometimes classifiable as dullness, often as feeblemindedness
Below 70	Definite feeblemindedness

The intelligence quotients presented in Table XXXVI to XXXVIII inclusive may be summarized into a single tabulation. This is presented in Table XL. The median of this table raises an important question. Its value is 1.06 which implies one of two things: either that the standards established for the different age groups have been set a little low or that Champaign represents a group slightly above the average in mentality. It is the opinion of the writer that the latter conclusion should be drawn.

An examination of Table XL gives considerable material for speculation. In the light of the percents shown in this distribution, it might be said that 1.9 percent of the school population in Champaign was definitely feeble-minded. Of course, a few pupils may have fallen into this group because they failed in the test through physical causes and not through lack of mental ability. But on the other hand, it might be added that there were several children in the schools who were too low in mentality to take the tests. Due to the fact that the teachers did not always give the data for these children, several of them were not included in the study. Some of them are hopeless imbeciles and probably will never learn to read or write.

If we adopt the standards of normality which are usually used, we shall find that only 46 percent of the school population is in the normal group—i.e., in the group which ranges from 0.90 to 0.110. The next higher group, 1.10 to 1.19 contains 23 percent. These are the two groups in which the greatest number of cases occur. Thus about seven out of every ten children in the Champaign schools appear to be either "normal" or "super-

ior," using Terman's classification. Below this large central group we have approximately 15 percent of the school population; above it, almost 17 percent.

TABLE XL. INTELLIGENCE QUOTIENT DISTRIBUTIONS FOR THE ENTIRE SCHOOL POPULATION

I. Q.	No. of Pupils	Percent
0.50–0.59	10	0.5
0.60–0.69	27	1.4
0.70–0.79	81	4.0
0.80–0.89	181	8.9
0.90–0.99	399	19.7
1.00–1.09	533	26.3
1.10–1.19	464	22.9
1.20–1.29	230	11.4
1.30–1.39	81	4.0
1.40–1.49	20	1.0
1.50–1.59	4	0.2
Total	2,030	
Median	1.06	

In its school administration Champaign should make an attempt to provide for these two extreme groups. At present they are found in the regular classes. There are but few attempts to meet the special needs of these children. The problem is not so complicated as this table might lead one to believe. The poorest 15 percent are found almost entirely in the grades below the eighth. Moreover, they are much more numerous in some sections of the city than in others. This fact will make it possible to provide special classes for them in which the curriculum can be modified to meet their particular needs. This group probably should be provided with more of the vocational and less of the academic subjects. It is probable for example, that a course in dishwashing, sweeping, cleaning, and in other simple household duties would be very beneficial to the girls who have intelligence quotients below nine-tenths. An examination of Table XXXVI shows that there are a sufficient number of these children to provide classes of economical size from the administrative point of view.

The 17 percent who are above the large central group should also be specially provided for. These are the children

who could make rapid progress through the school, if given the opportunity. It is decidedly unsatisfactory to give these children the same bill of fare as is provided for the average and the mediocre. In instruction one "exposure" of most topics is sufficient for them. A single reading of their lesson suffices, where the average or mediocre child must read it several times. From every point of view it is wasteful to keep these children in the same classes with the other children. It is not altogether satisfactory to give them rapid promotion from grade to grade, and thus to allow them to skip parts of the work. Consequently, the only sensible thing to do is to provide special classes for them.

Special classes, especially those for subnormal children, will no doubt meet with some opposition on the part of the parents, if the classes are established too abruptly and without the utmost tact. The present classification of the children in the city, which permits the transfer of a child from one school to another for administrative reasons, also permits a grouping of the children which will secure these desired results without any special advertising of the fact.

c. Age-grade intelligence-quotient groups—Regarding the child three facts not thus far combined in our tables are of special administrative importance. These are his chronological age, his grade in school, and his intelligence quotient. Manifestly, if a child is chronologically older than the normal age for the grade in which he is located and has an intelligence quotient that indicates approximate normality, the sensible thing to do is to promote him to the next higher grade, giving him extra attention so that he may meet the deficiencies in his scholarship preparation which result from this unusual progress. Champaign would experience no difficulty from the double promotion of such children, because the city has provided "opportunity classes" which all children who are maladjusted may attend in order that they may receive extra attention and make up work. These "opportunity classes" could provide for children who are now found to be out of place and do it without special effort.

The three facts which we have mentioned—age, grade, and intelligence quotient—may be presented in tabular form, thus furnishing a convenient means for discovering how many children may be considered improperly placed. Table XLI gives the distributions of intelligence quotients at each age for the the sixth grade and may serve as an example of this type of rep-

resentation. An examination of this table shows that comparatively few retarded children have an intelligence quotient high enough to warrant their promotion. Perhaps eleven of the 13 year-olds and one of the 14 year-olds might be advanced. But when we consider the fact that the intelligence level of the Champaign system is 1.06 it may be questioned whether it is

TABLE XLI. AGE-INTELLIGENCE-QUOTIENT DISTRIBUTION FOR SIXTH GRADE

I. Q.	Age in Years					
	10	11	12	13	14	15
0.50–0.59						
0.60–0.69					4	2
0.70–0.79			1	8	9	3
0.80–0.89		1	10	17	5	3
0.90–0.99	1	7	26	14	2	
1.00–1.09		20	26	8	1	
1.10–1.19	1	23	15	3		
1.20–1.29	1	11	1			
1.30–1.39	1	5				
1.40–1.49		1				
1.50–1.59	1					
Total	5	68	79	50	21	8
Median	1.25	1.13	1.01	0.90	0.77	0.77

advisable to promote any of these children except the one who is 14 years old and whose intelligence quotient is more than 1.00. In other words, there is comparatively little maladjustment from this point of view and certainly no extreme maladjustment. This table, which is typical of the other grades, emphasizes the fact that we shall always have retardation in our schools as long as children are grouped in classes where all must take the same curriculum without reference to their ability. It seems to the writer that the most sensible provision which can be made is to classify the children into at least three groups based on the intelligence quotient and then to prepare courses of study suited to these groups.

So far as ability to do higher-grade work is concerned it is not the older but the younger children who are really retarded. The reader will observe the consistently higher ranges of intelligence quotients among the younger children. The 15 twelve-year-old children whose intelligence quotients are between 1.10 and 1.19 have an average mental age of 13.8 years.

If they had received proper instruction and had been advanced according to their ability they would be in the seventh grade—some of them in the eighth. Similarly the eleven eleven-year-olds whose intelligence quotients are between 1.20 and 1.29 have mental ages which would entitle them to belong to the seventh grade. The five children of the same age who have intelligence quotients between 1.30 and 1.39 would be in the eighth grade, if their advancement had kept pace with their mental development. Whether such children should suddenly be promoted to the grade to which by mental age they belong is a debatable question. Such a belated adjustment would be at best a makeshift compared to the gradual adjustment that would have been possible if these children could have been identified early in their school career. Meanwhile, however, it is proper to point out that when pupils have been boldly promoted in accordance with their mental ages, even after they have been "discovered" relatively late in their school course, they have usually maintained themselves with credit in their advanced grades.

C—Summary and Comments

I. Schools vary appreciably in the distribution of pupils according to mentality. Some of the rooms in the same school are much superior to others. These large differences might have been anticipated from the general opinions of the teachers and supervisors.

II. Each child's departure from the median of his grade, in terms of the quartile deviation, serves as a convenient means of comparing his performance with that of another child whether the latter be in the same grade or not—or indeed whether he has taken the same test or not. This measure serves also to indicate the large irregularities in the classification of the children.

III. The intelligence quotient is a better means of measuring the individual variability of children within a room because it emphasizes the educational possibilities of each child. It thus becomes the best device, from the standpoint of the teacher, for measuring the "brightness" of the children.

IV. Grouping the children in Champaign on the basis of intelligence quotients, 69 percent of the school pupils may be considered of approximately uniform ability. Above this central homogeneous mass there are 17 percent who are above average and of high ability. Below it are 15 percent who are as inferior

as the others are superior. Special provisions with modified curriculums should be made for each of these two groups.

V. An age-intelligence-quotient table for each grade furnishes the best device for analyzing a school situation. This representation reveals the situation at a glance. If there are any maladjustments among the older groups, the fact is self-evident when the data are tabulated in this form. When the pupils who are out of place have been identified, it is a comparatively simple matter to apply the necessary remedies.

APPENDIX

I. *Incomplete Pictures as Tests*—The incomplete picture has been used consistently as a mental test since Binet introduced it in his scale. When it became necessary to devise a test for illiterates and foreigners in the army, incomplete pictures were accepted as a promising form of test. In the Beta examination devised by the army workers this form of test was used with adults of varying degrees of mental maturity, although the value of the incomplete picture as a test was comparatively unknown. Moreover, the data presented in Part II for the Virginia Delta I Scale raises the question of the suitability of the picture completion test for children beyond the primary grades because it shows almost no discriminative power for grades III to VIII. In the light of these facts it seems worth while to consider the value and possibilities of incomplete pictures as tests. Bearing upon this question are the results of a study of this form of test which has been made by the writer.

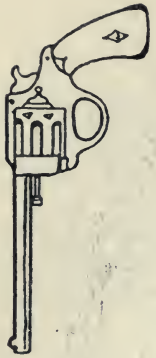
The data secured show that very few purely incomplete pictures are difficult enough to test the intelligence of normal children who are more than nine years of age. Furthermore, normal children of the primary grades are sometimes very much puzzled by incomplete pictures, although they may show a reasonable degree of keeness in other respects. In other words, the ability to recognize the omissions is a more or less specialized ability which depends to a certain extent on the type of experience the individual has had. When the problem of investigating the incomplete picture was first attacked it was fondly hoped that a series of incomplete pictures might be found which would present a range of difficulty capable of testing pupils from the primary through the grammar grades. This hope was not realized and the writer is led to the conclusion that incomplete pictures are of little value as tests of intelligence above the primary grades.

Thirty-eight incomplete pictures were submitted to approximately one thousand four hundred children. Of the total number of pictures presented twenty were found to be suitable for use in a test for the primary grades. These are presented in a "Picture Completion Test" published by the Bureau of Educational Research. Figure 1 shows this test.

FIGURE 1. PICTURE COMPLETION TEST BY DR. CHARLES E. HOLLEY

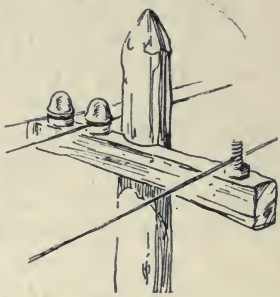
1 2 3 4 5 6 7 8 9

FIGURE 1—(*Continued*)

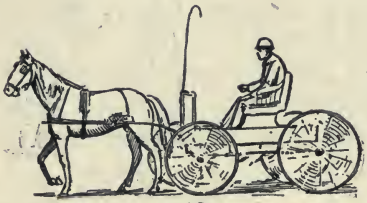


10

11



12

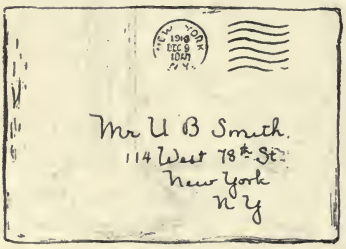


13

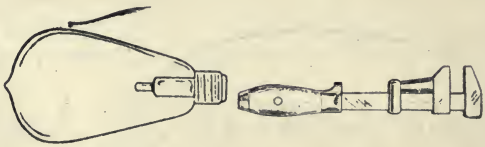
14

15

16



17



18 19

20

The difficulty in terms of the percent of correct responses was obtained for each of these twenty pictures. Table XLII shows these percents for the first three grades.

II. *Sex Differences*—The data gathered in Champaign from all the tests were examined from the standpoint of the difference between boys and girls. The results were such as to lead one to conclude that there are no real sex differences in general intelligence which may be revealed by these general

TABLE XLII. PICTURE COMPLETION TEST. PERCENTS CORRECT FOR EACH PICTURE BY GRADES

Number	Grade		
	I	II	III
1	91	100	100
2	86	95	97
3	84	98	97
4	77	89	97
5	77	89	95
6	63	82	94
7	60	82	86
8	56	62	87
9	49	59	69
10	46	56	68
11	47	56	61
12	44	63	78
13	44	61	72
14	39	59	78
15	30	37	53
16	30	33	53
17	25	39	66
18	21	72	91
19	15	45	51
20	10	30	53

tests. The medians and averages for the two sexes were approximately the same in the various tests, all differences being small enough to justify one in attributing them to chance factors.

III. *Administration by Teachers*—It is highly desirable to have intelligence scales that can be administered by the individual teacher. If, however, several rooms are to be compared, it is better, in practice, for a supervisor to give the tests than for the individual teacher to do so. Although most teachers will do their best to follow instructions accurately and thus to secure uniform results, a small minority will persist in varying conditions to suit their own ends. From many points of view, it will

be better if the pupils take intelligence tests under the direction of supervisors or of persons other than the room teachers. The results will be comparable from room to room and they may be made the basis of administrative measures in a way that would not be the case if their reliability were in doubt. Since nearly every test requires practice for its successful administration, the supervisor who administers a test several times becomes practiced in its details and thus secures the complete cooperation of the pupils. With the more difficult scales the supervisor can take the time to perfect his method by practicing on his friends before he administers the scale in the classroom where the results are important. Individual teachers cannot spare the time and trouble needed to perfect their technique even if they are entirely in sympathy with the work. Consequently, it is much better if all tests of this nature are administered by supervisors.

IV. *Scoring*—If mental tests are to be used in a way that will contribute most to school problems they must be scored very accurately. Where a child's future is to be influenced by the result it is vitally important that his score be as nearly correct as the test will permit.

The best results are secured if the scoring is done by a few careful workers who have been trained for their duties instead of by a larger number of people who devote only a little time individually to the work. It requires much valuable time to instruct the scorers in the methods of evaluating the various parts of the pupil's answers, and as a rule, several days of practice are needed before they can become proficient. Consequently, paid trained workers are a decided economy over volunteer workers. As far as possible nothing should be left to independent judgment. Instructions should be prepared which will cover every possible case. The system of indicating rights and wrongs on the tests should be worked out so carefully that it will economize time and eliminate chances for error. All of these details should be covered in the training of the clerical workers.

It is possible to secure good clerical workers among the student body of the average high school and college. Scoring takes good eyesight and ability to learn quickly. People who score for a few days on a test acquire increased proficiency from day to day and reach their maximum in about a week. This fact suggests that it is uneconomical to employ a large number of workers for a small project. It is better to secure a few capable

workers and allow them to perform the same type of work for a longer time than would be needed for the larger group. They will become very skilled and will do the work at a much smaller unit cost.

It is best to rescore or check nearly every operation that is involved in rating a set of papers. This checking should be done by persons other than those who scored the papers the first time. A second scoring will catch most of the errors.

Stencils facilitate the work of scoring by economizing eye movements and pencil marks. Some types of material lend themselves to the use of transparent stencils. These stencils may be made from the celluloid used as window material for automobile curtains. Transparent paper may also be used—especially when durability is not necessary. Ink dots or lines may be so placed on this material that they will coincide with the marks that the pupils must make in indicating the right answers to the tests. In other types of tests cardboard stencils may be made which will enable the scorers to check the answers quickly. Every device should be employed which will economize time and insure a high degree of accuracy. Any device which leads to simplicity removes sources of error.